

**MODELO ESTADÍSTICO PARA LA  
LOCALIZACIÓN DE FALLAS EN SISTEMAS DE  
DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA**

**Ing. JORGE ANDRÉS CORMANE ANGARITA**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**Facultad de Ingenierías Físico Mecánicas**

**Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones**

**Bucaramanga**

**2006**

# MODELO ESTADÍSTICO PARA LA LOCALIZACIÓN DE FALLAS EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA



Trabajo de investigación para optar al título de Magíster en Ingeniería  
Eléctrica

Autor: Ing. Jorge Andrés Cormane Angarita

Director: Prof. Dr. Hermann Raúl Vargas Torres

Codirector: Prof. Dr. Gabriel Ordóñez Plata

**UNIVERSIDAD INDUSTRIAL DE SANTANDER**

**Facultad de Ingenierías Físico Mecánicas**

**Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones**

Bucaramanga

2006

## **TITULO\***

MODELO ESTADÍSTICO PARA LA LOCALIZACIÓN DE FALLAS EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA

## **AUTOR\*\***

JORGE ANDRÉS CORMANE ANGARITA

## **PALABRAS CLAVES**

Caídas de tensión, Calidad de potencia, Clasificación, Índices de calidad, Localización de fallas, Mezclas finitas, Modelo estadístico, Modelos de mezclas.

## **DESCRIPCIÓN**

El restablecimiento del servicio de energía eléctrica ante interrupciones está directamente relacionado con la atención a fallas por parte de los operadores de red. Dicho problema se trata bajo la óptica de la localización de fallas en el sistema y la reducción del tiempo de atención a las mismas. Mejorar la confiabilidad de las redes de distribución requiere de una gran inversión, pero no todas las empresas están en capacidad asumirlas. Por lo tanto cualquier estrategia que permita mejorar la confiabilidad, se debe reflejar directamente en la disminución de la duración y frecuencia de las interrupciones y mas concretamente en los índices DES y FES.

En este artículo se presenta una alternativa de solución a los problemas de continuidad del suministro asociados a la localización de fallas, con el objeto de disminuir el tiempo de restablecimiento del sistema. Se propone una metodología de tipo estadístico fundamentada en la estimación de la localización del punto de falla a partir del modelado estadístico de los eventos ocurridos en la red utilizando la técnica de mezclas finitas. La metodología se implementa en una herramienta *software* que emplea para la construcción del modelo, la extracción de patrones característicos de las señales registradas por los equipos de medida junto con los parámetros y topología propios de la red durante el evento. La utilización de esta metodología tiene por objeto brindar una alternativa económica y de fácil implementación para el desarrollo de estrategias orientadas a mejorar la confiabilidad a partir de la disminución de los tiempos de atención y recuperación del sistema.

\*Trabajo de Investigación o Tesis.

\*\*Facultad de Ingenierías Físico Mecánicas, Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones, Hermann Raúl Vargas Torres.

**TITLE\***

STATISTICAL MODEL FOR FAULT LOCATION IN DISTRIBUTION SYSTEMS

**AUTHOR\*\***

JORGE ANDRÉS CORMANE ANGARITA

**KEY WORDS**

Fault Location, Power Quality, Sags, Quality of Service Indices, Statistical Models, Classification, Finite Mixture Model, Densities Mixture Model.

**DESCRIPTION**

Electric power service restoration upon the occurrence of interruptions is directly related to the attention of faults by system operators. Such a problem is faced from the viewpoint of fault location and reductions in the time of attention. Improving the reliability of distribution networks requires a great investment that most utilities are not willing to make. Therefore, any strategy designed to improve reliability should be directly reflected in the decrease of duration and frequency of interruptions, particularly in the improvement of DES and FES indices.

In this article an alternative to the solution of continuity problems associated to fault location is presented. The purpose is to diminish the time of system restoration. A statistical methodology is proposed based on the estimation of fault location using statistical modeling of events in the network utilizing the technique of finite mixtures. The methodology is implemented in a software tool that employs characteristic patterns signals registered by meters along with parameters and network topology for the construction of the model. The utilization of this methodology has the aim of offering an alternative both economically feasible and easily implementable for the development of strategies oriented to the improvement of reliability obtained from the decrease in the times of attention and system recovery.

\*Trabajo de Investigación o Tesis.

\*\*Facultad de Ingenierías Físico Mecánicas, Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones, Hermann Raúl Vargas Torres.

# AGRADECIMIENTOS

Quiero agradecer a las personas que colaboraron directamente en la consecución de este trabajo, en especial a los ingenieros Libardo Villamizar, Carlos Quiñones, Edwin Salamanca, Marcos Suárez, Luis López, Angélica Romero y Diego Beltrán, los cuales asumieron el compromiso al conformar mi grupo de trabajo.

A los profesores Hermann Raúl Vargas Torres, Gabriel Ordóñez Plata y Gilberto Carrillo Caicedo por su orientación y asistencia.

A mis compañeros, los ingenieros Víctor Barrera, Juan Carlos Suárez, Carlos Rodríguez, Carlos Cárdenas, Valdomiro Vega, Manuel Ortiz y Wilson Giraldo.

*A mis padres Álvaro David e Inés Elvira*  
*A mis hermanos Álvaro Rafael y Juan David*  
*A mis tíos Carlos Eduardo y Ana Helena*  
*A mi novia Claudia*

*“Yo soy artífice y protagonista de mi propia historia”*

Francisco Bedoya Bedoya

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	2
1.2. Objetivos del trabajo de investigación . . . . .	3
1.3. Estructura del documento . . . . .	4
<b>2. Localización de fallas: ópticas determinística y estadística</b>	<b>7</b>
2.1. Introducción . . . . .	7
2.2. Descripción de un sistema típico de distribución . . . . .	7
2.3. Fallas de cortocircuito . . . . .	9
2.4. Antecedentes en la localización de fallas . . . . .	10
2.5. Óptica determinista (Técnicas basadas en el modelo) . . . . .	10
2.6. Óptica estadística (Técnicas basadas en el conocimiento) . . . . .	12
2.6.1. Técnicas Estadísticas (Reconocimiento de Patrones) . . . . .	14
<b>3. Análisis estadístico multivariado</b>	<b>19</b>
3.1. Introducción . . . . .	19
3.2. Definición . . . . .	19
3.3. Conceptos básicos utilizados . . . . .	20
3.3.1. La matriz de datos . . . . .	20
3.3.2. El valor teórico . . . . .	20
3.3.3. Vector de medias $\bar{\mathbf{x}}$ . . . . .	20
3.3.4. Matriz de varianzas y covarianzas $\mathbf{S}$ . . . . .	21
3.3.5. Matriz de correlación $\mathbf{R}$ . . . . .	21
3.3.6. La distancia . . . . .	22
3.4. Supuestos utilizados . . . . .	24
3.4.1. Normalidad . . . . .	24
3.4.2. Homocedasticidad . . . . .	26
3.4.3. Linealidad . . . . .	26

3.5. Planteamiento inicial . . . . .	27
3.6. Análisis de componentes principales - ACP . . . . .	27
3.6.1. Cálculo de las componentes principales . . . . .	28
3.6.2. Selección del número e interpretación de las CP . . . . .	28
3.7. Método de las K-medias . . . . .	29
3.8. Mezclas de distribuciones - MD . . . . .	30
<b>4. Metodología para la construcción del modelo estadístico</b>	<b>35</b>
4.1. Introducción . . . . .	35
4.2. Sistema prototipo . . . . .	36
4.3. Simulación del sistema . . . . .	37
4.4. Caracterización del sistema . . . . .	39
4.4.1. Descriptores en estado estable . . . . .	40
4.4.2. Descriptores en estado transitorio . . . . .	41
4.5. Análisis de los datos . . . . .	44
4.6. Construcción del modelo de mezclas . . . . .	46
4.6.1. Estructura del modelo . . . . .	46
4.6.2. Validación del modelo . . . . .	48
<b>5. Pruebas y Resultados</b>	<b>51</b>
5.1. Introducción . . . . .	51
5.2. Determinación del sistema y los datos a utilizar . . . . .	51
5.3. Entrenamiento . . . . .	53
5.4. Validación . . . . .	59
5.5. Aplicación del algoritmo Ratan Das . . . . .	61
5.6. Aplicación del híbrido Ratan Das - Modelo de Mezclas . . . . .	65
<b>6. Estimación de los índices de calidad del servicio DES y FES</b>	<b>71</b>
6.1. Introducción . . . . .	71
6.2. Índices de calidad del servicio . . . . .	72
6.3. Determinación de los tiempos de indisponibilidad . . . . .	72
6.4. Método para la estimación de los índices de calidad . . . . .	74
6.4.1. Determinación del estado de los tramos . . . . .	74
6.4.2. Cálculo del tiempo de interrupción de los tramos . . . . .	75
6.4.3. Cálculo de la indisponibilidad por tramo . . . . .	76
6.5. Impacto de la localización de fallas en los índices DES y FES . . . . .	78

<b>Índice general</b>	<b>III</b>
<hr/>	
<b>7. Conclusiones y trabajos futuros</b>	<b>83</b>
7.1. Conclusiones . . . . .	83
7.2. Aportes . . . . .	85
7.3. Trabajos Futuros . . . . .	86
<b>A. Parámetros del sistema de distribución de prueba</b>	<b>89</b>
<b>Bibliografía</b>	<b>93</b>



# Índice de figuras

2.1. Sistema típico de distribución. . . . .	8
2.2. Diagrama de conglomerados que muestra la variación dentro y entre conglomerados. . . . .	15
3.1. Funciones de densidad normal univariada. . . . .	24
3.2. Función de densidad normal multivariada. (a) Distribución normal bivariada. (b) Familia de curvas asociada. . . . .	25
3.3. Gráfica de dispersión con relaciones de homocedasticidad y heterocedasticidad. (a) Homocedasticidad. (b) Heterocedasticidad . . . . .	26
4.1. Diagrama unifilar del sistema de distribución. . . . .	36
4.2. Sistema de distribución implementado en ATP/EMTP. . . . .	37
4.3. Señal monofásica de tensión durante una falla. (a) Forma de onda. (b) Valor eficaz. . . . .	38
4.4. Descriptores monofásicos. (a) Descriptores de tensión. (b) Descriptores de corriente. . . . .	40
4.5. Tratamiento de la señal transitoria. (a) Coeficientes de detalle (TW). (b) Componentes de frecuencia (FFT). . . . .	42
4.6. Procedimiento para la extracción de descriptores de estado transitorio. . . . .	44
4.7. Resultados del ACP en los descriptores para fallas Monofásicas. (a) Análisis de correlación multivariado. (b) Porcentaje de información en cada CP. . . . .	45
4.8. Resultados del ACP en los descriptores para fallas Bifásicas a tierra. (a) Análisis de correlación multivariado. (b) Porcentaje de información en cada CP. . . . .	45
4.9. Diagrama de flujo de la construcción y entrenamiento del modelo estadístico. . . . .	47
4.10. Diagrama de flujo de la validación y uso del modelo estadístico. . . . .	49
5.1. Zonas de falla del sistema. (a) Esquema para zonas monofásicas. (b) Esquema para zonas bifásicas y trifásicas. . . . .	54
5.2. Distribución de los datos de entrenamiento falla monofásica. . . . .	55
5.3. Distribución de los datos de entrenamiento falla bifásica. . . . .	55
5.4. Distribución de los datos de entrenamiento falla bifásica a tierra. . . . .	56

5.5. Distribución de los grupos según rangos de resistencia de falla para una falla monofásica (EC2). . . . .	57
5.6. Distribución de los grupos (zonas) asociados al grupo 1 (rango 1) de la etapa EC3. . .	58
5.7. Resultados por Tipo de falla . . . . .	61
5.8. Resultados por Tipo de falla y fase Fallada . . . . .	62
5.9. Cantidad de aciertos por nodo y por tipo de falla . . . . .	65
5.10. Resultados del modelo híbrido. (a) Caso 1. (b) Caso 2. (c) Caso 3. (d) Caso 4. . . .	66
5.11. Resultados del modelo híbrido para los tipos de falla evaluados . . . . .	66
5.12. Resultados del modelo híbrido para cada nodo del sistema . . . . .	67
5.13. Porcentaje de los resultados distribuidos por caso . . . . .	68
5.14. Comparativo entre las estimaciones realizadas por el algoritmo Ratan Das y las reducidas por el modelo estadístico . . . . .	69
6.1. Intervalos del tiempo total de interrupción. . . . .	72
6.2. Circuito fallado. . . . .	79
6.3. Localización de fallas (métodos convencionales). . . . .	79
6.4. Localización de fallas (métodos algorítmicos). . . . .	80
6.5. Localización de fallas (técnicas de clasificación). . . . .	80
6.6. Localización de fallas (métodos algorítmicos y técnicas de clasificación). . . . .	81
A.1. Diagrama unifilar del sistema de distribución. . . . .	89

# Índice de tablas

2.1. Comparación de los algoritmos de localización de fallas para sistemas de distribución	11
4.1. Simulaciones del sistema de distribución en ATP/EMTP	37
4.2. Descriptores de tensión y corriente	41
4.3. Características de la TW y la FFT	43
5.1. Valores de resistencia de falla para el proceso de entrenamiento y validación	52
5.2. Número de simulaciones realizadas para cada proceso	52
5.3. Barras que conforman cada una de las zonas de falla	53
5.4. Resultados de validación (Etapa EC1)	59
5.5. Resultados de validación (Etapa EC2)	60
5.6. Resultados de validación (Etapa EC3)	60
5.7. Ejemplo de clasificación (Etapa EC3)	61
5.8. Resultados por Tipo de Falla	62
5.9. Resultados por Tipo de Falla y Fase	63
5.10. Resultados por Nodos	64
6.1. Matriz de estados	75
6.2. Tiempos de interrupción	75
6.3. Matriz de tiempos de interrupción por tramos	76
6.4. Interrupciones aportadas según estado	77
6.5. Matriz de interrupciones	77
6.6. Matriz de indisponibilidades	78
A.1. Parámetros del generador.	90
A.2. Parámetros de las de cargas del sistema.	90
A.3. Parámetros de las líneas	91
A.4. Factor de potencia de las cargas.	91



# Capítulo 1

## Introducción

Desde principios de los noventa los temas concernientes a la calidad del servicio han estado en constante discusión a nivel internacional <sup>1</sup> [Jaimes et al., 2004]. Esto se debe al ambiente de libre competencia del sector eléctrico, donde la calidad es un buen reflejo de la participación de las empresas en el mercado energético y uno de los aspectos de mayor interés frente a las exigencias del sector a nivel mundial [Bollen, 2000].

En Colombia, a partir del establecimiento del nuevo marco regulatorio del sector eléctrico, se pretende fortalecer la distribución y el mercado minorista desde la óptica de la calidad del producto <sup>2</sup> [CREG, 2006], [Jaimes et al., 2004]. La dependencia existente entre el ser humano y la energía eléctrica, ha determinado que cada vez y con más insistencia se requiera que el suministro se realice bajo criterios de seguridad, confiabilidad y calidad [Cormane et al., 2004]. Tales acciones se han promovido por medio de tarifas fijas en los usuarios finales y a través de la aplicación de estándares para la prestación del servicio por parte de las empresas del sector <sup>3</sup> [Carrillo y Ordóñez, 2002].

La continuidad en el suministro es uno de los aspectos más importantes para el cliente. Dicha importancia se origina por el impacto económico y social generado a partir de la ocurrencia de interrupciones [Bollen, 2000], [Peñuela et al., 2004]. Aunque económicamente no es viable lograr un 100% de confiabilidad, las empresas se están esforzando en solucionar el problema de la continuidad mediante una adecuada planificación y operación del sistema [Jaimes, 2003].

---

<sup>1</sup>La calidad del servicio prestado enmarca aspectos concernientes a la continuidad del suministro.

<sup>2</sup>La calidad del producto, de onda o de potencia suministrada, se refiere a las perturbaciones y variaciones del estado estacionario de tensión y corriente, regulación de frecuencia y tensión, armónicos, fluctuaciones de tensión y factor de potencia.

<sup>3</sup>Empresas distribuidoras de energía eléctrica y operadores de red.

En la actualidad, se estima que el 80% de las interrupciones son causadas por fallas en sistemas de distribución. Pensar en la implementación de algoritmos similares a los utilizados en los sistemas de transmisión no es viable en sistemas de distribución puesto que su funcionamiento se ve afectado por las características técnicas (topología, secciones no homogéneas, múltiples derivaciones, equipos de medida) y operativas (operación radial), comprometiendo la eficiencia en el desempeño del mismo [Bollen, 2000], [Das, 1998].

Existe variedad de métodos para la localización de fallas en sistemas de distribución. Estos métodos se clasifican en tres categorías. La primera, métodos que detectan componentes de alta frecuencia de ondas viajeras. La segunda, métodos que utilizan el cálculo de la impedancia y la tercera, métodos de inspección visual, que implican el patrullaje y reconocimiento de la línea [Das, 1998], [Zhu et al., 1997].

La Universidad Industrial de Santander a través del Grupo de Investigación en Sistemas de Energía Eléctrica GISEL está interesada en desarrollar trabajos de investigación en el tema de la localización de fallas en los sistemas de distribución de energía eléctrica. Tema directamente relacionado con el área de monitorización de la continuidad del suministro e índices de calidad del servicio, los cuales, son tópicos de actualidad científica que conciernen a la línea de investigación de la calidad del suministro de energía eléctrica en la que el grupo GISEL posee amplia trayectoria y experiencia por parte de sus investigadores.

## 1.1. Planteamiento del problema

Los nuevos esquemas regulatorios del sector eléctrico, han estimulado a que muchas de las empresas aborden la problemática planteada a partir de la reducción del número y la duración de interrupciones, con el fin de fomentar la competitividad y la calidad en la prestación del servicio. La idea central de este trabajo es proponer una alternativa de solución a los problemas de continuidad del suministro asociados al problema de la localización de fallas. Este radica fundamentalmente en la dificultad que presentan las empresas distribuidoras de energía en la implementación de algoritmos y esquemas de última tecnología para la identificación y localización de fallas en sus redes, dadas las implicaciones técnicas, económicas y operativas requeridas.

Para la consecución de esta meta se propuso disminuir el tiempo de restablecimiento de las condiciones de operación normal del sistema [Girgis et al., 1993]. Esta situación permite

mejorar la confiabilidad a partir del planteamiento de estrategias orientadas a la disminución de los tiempos de atención y recuperación del sistema, brindando una alternativa de bajo costo y fácil implementación. Como parte del desarrollo del trabajo se planteó aprovechar los equipos, la información, el personal y la experiencia, para mejorar la prevención y atención oportuna de fallas.

Dicha propuesta se apoya en las últimas modificaciones a las normas de calidad de la potencia eléctrica realizadas por la CREG <sup>4</sup>, respecto a la monitorización de la continuidad del suministro y calidad de la potencia suministrada [CREG, 2006]. Es de notar que los resultados mostrados en el estudio fueron obtenidos utilizando únicamente datos de simulación. El siguiente paso luego de la culminación de este estudio, es la aplicación del modelo a datos reales (mediciones).

El desarrollo del trabajo de investigación está justificado en tres tópicos esenciales, el primero, relacionado con la necesidad actual de mejorar los niveles de calidad por parte de las empresas del sector, el segundo, en las entidades educativas interesadas; la Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones de la Universidad Industrial de Santander, y la Facultad de Ingeniería Eléctrica de la Universidad Tecnológica de Pereira, cuyo interés se orienta al desarrollo y divulgación de líneas de investigación enfocadas a la resolución de problemas de actualidad científica, tal como lo es el problema de localización de fallas en sistemas de distribución, y tercero, por parte de COLCIENCIAS e ISA S.A E.S.P interesados en fomentar e incentivar el desarrollo y la innovación tecnológica en Colombia.

## **1.2. Objetivos del trabajo de investigación**

El objetivo general de este trabajo de investigación es el desarrollo de un método alternativo para la localización de fallas en sistemas de distribución de energía eléctrica a partir de un modelo estadístico que contribuya al mejoramiento de la calidad del servicio en las empresas del sector.

El modelo está basado en la utilización de la técnica mezclas finitas<sup>5</sup>, como clasificador estadístico, buscando innovar debido a las pocas aplicaciones en los temas relacionados con el área de interés. La idea de emplear este clasificador es con el fin de aprovechar de mejor manera la información disponible, proporcionada por los equipos de medición y control, para la rápida ubicación de la falla.

---

<sup>4</sup>Comisión Reguladora de Energía y Gas - CREG.

<sup>5</sup>Llamada también mezclas de distribuciones.

A la vista del objetivo general expuesto anteriormente se han trazado los siguientes objetivos específicos:

1. Establecer patrones para la identificación y caracterización de fallas con base en la información del comportamiento del sistema de distribución ante la ocurrencia de una falla.
2. Analizar comparativamente el comportamiento en estado estable y transitorio de los sistemas de distribución de energía ante la ocurrencia de una perturbación que altere las condiciones de operación normal.
3. Proponer un modelo estadístico para la localización de fallas en sistemas de distribución de energía eléctrica, basado en la aplicación de la técnica estadística Mezclas Finitas.
4. Formular estrategias que permitan mejorar la calidad del servicio a partir de información asociada a la localización de fallas.

### 1.3. Estructura del documento

Este documento se encuentra organizado en 8 capítulos y un apéndice. El capítulo 2 presenta la problemática de la localización de fallas desde dos ópticas, la primera, determinista relacionada con los métodos algorítmicos y la segunda, estadística relacionada con técnicas de clasificación de datos multivariados. Se destacan algunos de los trabajos realizados en relación con el tema y se presenta el comparativo entre algunos métodos algorítmicos consultados.

El capítulo 3 introduce la formulación matemática correspondiente e ilustra acerca de la operación y requerimientos de los algoritmos empleados en la construcción del modelo.

El capítulo 4 presenta la metodología propuesta para la construcción del modelo. Se describe detalladamente cada una de las etapas planteadas durante la investigación. Adicionalmente se analiza comparativamente el comportamiento en estado estable y transitorio del sistema de distribución de energía ante la ocurrencia de una falla.

El capítulo 5 presenta y discute los resultados obtenidos al aplicar el modelo desarrollado en este trabajo de investigación. Se describe la información utilizada en las pruebas realizadas.

El capítulo 6 analiza el impacto del modelo sobre la calidad del servicio. Se presentan y discuten los resultados acerca de las mejoras potenciales obtenidas a partir de información de localización de fallas.

Finalmente, en el capítulo 7 se exponen las conclusiones y aportes realizados con este trabajo. Se sugieren líneas de trabajo para desarrollos futuros.

Complementario a la estructura anterior se encuentra el apéndice A en el cual se exponen algunos parámetros que se tuvieron presente en el circuito de simulación.



# Capítulo 2

## Localización de fallas: ópticas determinística y estadística

### 2.1. Introducción

Este capítulo enmarca la problemática de la localización de fallas en sistemas de distribución de energía eléctrica en el contexto del trabajo de investigación y revisa el estado actual del tema.

El tema de localización de fallas en sistemas eléctricos de transporte de energía eléctrica sistemas de transmisión y distribución ha sido de considerable interés para el sector eléctrico, tanto para ingenieros como investigadores durante más de 30 años. El incremento en las investigaciones en este tema ha desembocado en el desarrollo de métodos novedosos y cantidades de publicaciones en revistas técnicas de impacto internacional. Cabe resaltar que en su mayoría, las investigaciones han estado encaminadas a la localización de fallas en sistemas de transmisión, debido, básicamente al impacto socioeconómico generado sobre el sistema de potencia en general.

### 2.2. Descripción de un sistema típico de distribución

Se entiende por sistema de distribución de energía eléctrica a la disposición adoptada por los elementos del sistema tales como conductores, transformadores, protecciones y consumidores (residenciales, comerciales e industriales), con el fin de lograr que la energía generada por las centrales eléctricas y transportada a través del sistema de transmisión, pueda ser utilizada en los sitios de consumo.

Los transformadores de distribución están conectados por el primario al sistema de Media

Tensión<sup>1</sup> y por el secundario a la red de Baja Tensión<sup>2</sup>, llevando la potencia a niveles de tensión de utilización para los usuarios [CIDET, 2005]. Las líneas de MT están conformadas principalmente por secciones trifásicas de las cuales se desprenden secciones bifásicas y monofásicas [Brown, 2002].

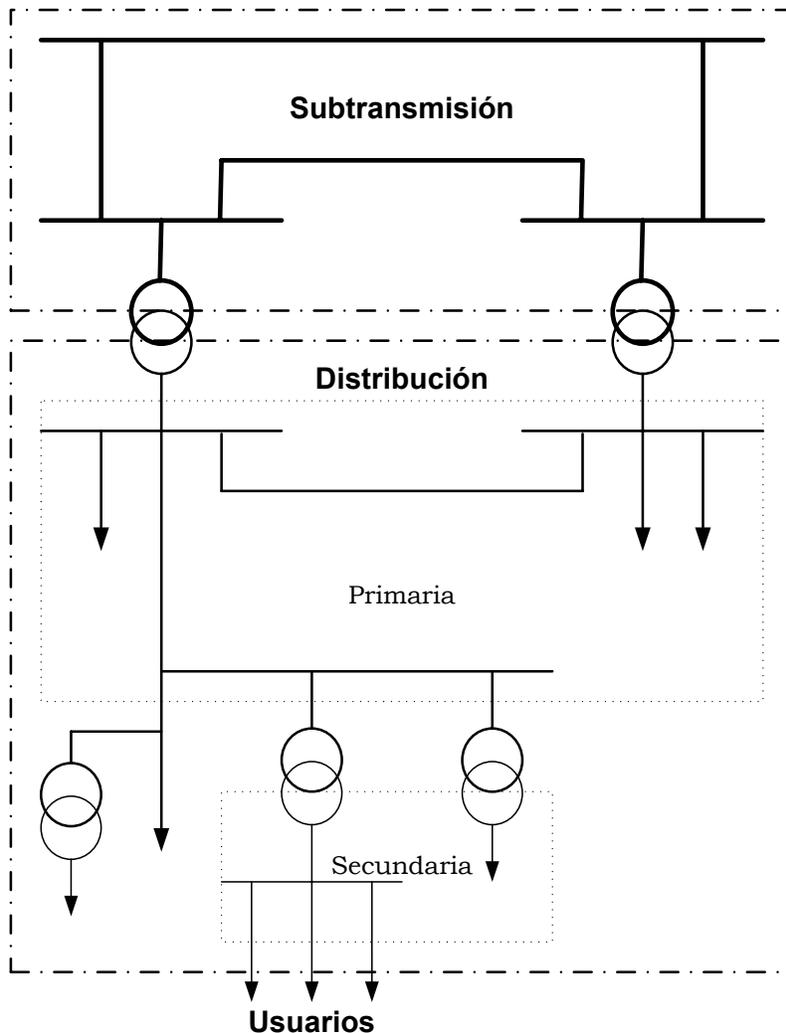


Figura 2.1: Sistema típico de distribución.

Los sistemas de distribución normalmente se operan de forma radial. Las cargas se encuentran conectadas y distribuidas de forma monofásica, bifásica y trifásica a lo largo de sus redes. Adicionalmente, su topología presenta secciones de línea de diferentes calibres (No-Homogéneas) debido a que las redes de distribución están sometidas a constantes ampliaciones originadas por el crecimiento del sistema y el incremento de la demanda [Das, 1998]. A continuación se realiza la descripción de un sistema de distribución típico mostrado en la figura 2.1.

<sup>1</sup>Niveles de utilización primarios  $1kV < V < 57,5kV$ .

<sup>2</sup>Niveles de utilización secundarios  $V < 1kV$ .

En la figura 2.1 se muestra un sistema típico de distribución el cual incluye circuitos de subtransmisión, subestaciones, alimentadores, transformadores, circuitos de distribución primaria y secundaria. Los circuitos de subtransmisión operan a tensiones que oscilan entre 115kV y 220kV, transportando grandes cantidades de energía a las subestaciones de distribución. Una subestación de distribución incluye transformadores de potencia, barrajes, líneas, reactores, condensadores, aisladores, interruptores y seccionadores. Los transformadores reducen la tensión de subtransmisión a niveles menores 13,2kV y 34,5kV para la distribución local.

## 2.3. Fallas de cortocircuito

En las interrupciones del suministro intervienen dos factores importantes, el primero, la ocurrencia de fallas al interior del sistema de distribución, y el segundo ocasionado por la operación de las protecciones. Habitualmente los sistemas de distribución experimentan fallas de cortocircuito. Estas se clasifican en cuatro categorías:

- Monofásica a tierra-LT
- Bifásica-LL
- Bifásica a tierra-LLT
- Trifásica-3LT.
- Trifásica a tierra-3LT.

Generalmente, entre el 70 % y el 80 % de las fallas presentes en los sistemas de distribución son monofásicas [Grainger y Stevenson, 1999]. Todas las fallas excepto las trifásicas causan desbalance en el sistema. La utilización de la teoría de las componentes simétricas facilita el análisis de las mismas [Anderson, 1995].

Las fallas nombradas se originan principalmente por defectos en los equipos del sistema, falta de mantenimiento de las servidumbres (presencia de animales y árboles), condiciones climáticas severas y factores humanos (vandalismo y errores operativos), manifestándose como sobrecargas en los equipos, variaciones en la frecuencia, disparos incorrectos de las protecciones y acciones operativas erróneas generando interrupciones en el sistema. Algunas perturbaciones pueden afectar a los usuarios, pero la correcta operación de las protecciones evita que estas se conviertan en interrupciones que desencadenen cortes del suministro y daños a los equipos eléctricos [Bollen, 2000], [Brown, 2002].

## 2.4. Antecedentes en la localización de fallas

Se han desarrollado diversos métodos para la localización de fallas. Uno de los métodos más empleados es la inspección visual de la línea. Éste método tiene el inconveniente de que se debe realizar a pie o en automóvil y apoyado por el uso de binoculares, ocasionando demoras en la localización e inconvenientes debidos al estado del clima. Los métodos existentes para la localización de fallas se han dividido clásicamente en tres categorías.

- **Fenómenos de Ondas Viajeras:** Se basa en el reconocimiento del comportamiento de ondas viajeras de tensión y corriente en condiciones de falla. Son métodos con alto grado de precisión y de complejidad, lo cual dificulta la implementación.
- **Componentes de Alta Frecuencia:** Se basa en la obtención de las componentes de alta frecuencia de señales de tensión y corriente en condiciones de falla. El inconveniente de éste método son los altos requerimientos de hardware y software.
- **Componentes de Frecuencia Fundamental:** Emplea las señales de tensión y corriente a frecuencia fundamental. Calcula la impedancia vista desde los terminales de la línea y la compara con los parámetros de la misma para estimar una distancia desde los terminales de la línea hasta el lugar de la falla.

Otro método utilizado se basa en indicadores ubicados en cada sección de la línea provistos de enlaces de comunicación por radio. Este método se hace poco viable a medida que el sistema de distribución se hace más extenso.

## 2.5. Óptica determinista (Técnicas basadas en el modelo)

Los sistemas de distribución son la principal fuente de energía para gran cantidad de usuarios industriales, comerciales y residenciales. Una falla al interior del sistema representa una interrupción del suministro y esta a su vez se traduce en pérdidas económicas por paradas en los procesos, pérdidas de información y daños en maquinarias e insumos. En la mayoría de los casos, las fallas eléctricas causan daños mecánicos los cuales se deben remediar de forma rápida y oportuna con el fin de restaurar la operación normal del sistema.

La restauración del suministro se puede facilitar si se conoce o se estima apropiadamente la ubicación de la falla, lo cual, es una de las principales razones por las que se hace necesario investigar a cerca de la localización de fallas en sistemas eléctricos, especialmente en circuitos

altamente ramificados, con diferentes calibres de conductores y geográficamente dispersos como son los de distribución.

El aumento del tamaño y la complejidad de los sistemas de distribución ha sido producto del notorio crecimiento de la demanda junto a las nuevas estrategias de mercado en el sector eléctrico. Lo anterior va ligado al compromiso de las empresas por mantener adecuados niveles de confiabilidad, reduciendo el tiempo de indisponibilidad de las redes debido a la ocurrencia de fallas. En consecuencia, diversas investigaciones han sido desarrolladas con el fin de realizar una estimación de la localización del punto bajo falla, en su mayoría orientadas a los sistemas de transmisión. Sin embargo, otros desarrollos se han centrado en los sistemas de distribución. Dichos trabajos se basan en el modelo del sistema y utilizan como variables de entrada, los valores de la componente fundamental de tensión y corriente, en prefalla y posfalla. No obstante, se presentan inconvenientes debido a las características propias de los circuitos de distribución, lo que origina problemas de múltiple estimación que comprometen la eficiencia y precisión del algoritmo. Se resalta que para el desarrollo de los algoritmos de localización, es necesario un conocimiento pleno de las características del sistema. Los principales estudios sobre el particular son reseñados brevemente a continuación en la tabla 2.1 [Cormane et al., 2005].

Tabla 2.1: Comparación de los algoritmos de localización de fallas para sistemas de distribución

ALGORITMOS		CARACTERÍSTICAS DEL SISTEMA DE DISTRIBUCIÓN				
Autores Consultados	Información Utilizada	Modelo Línea	Cargas Laterales	Heterogeneidad de la red	Tipo de Falla	Múltiple Estimación
RatanDas	Pre-Post	Larga	si	si	si	si
Srinivasan	Pre-Post	Larga	no	si	si	si
Girgis	Pre-Post	Corta	si	si	si	si
Myeon	Pre-Post	Corta	no	si	no	si
Yang	Pre-Post	Corta	no	si	si	si
Saha	Pre-Post	Corta	si	si	si	si
Aggarwal	Pre-Post	Corta	si	si	si	si
Warrington	Pre-Post	Corta	no	no	no	si
Takagi	Pre-Post	Corta	si	no	no	si
Novosel	Pre-Post	Corta	no	no	si	si

En la tabla 2.1 se presenta la comparación entre los algoritmos de localización de fallas

consultados. El motivo de la comparación es saber si los algoritmos contemplan en su estructura de cálculo las características típicas de los sistemas de distribución de energía eléctrica. Adicionalmente se busca conocer que tipo de información es necesaria para la óptima ejecución de los algoritmos, debido a que muchas veces las empresas no poseen en su infraestructura los medios necesarios para la determinación de las variables de entrada de los algoritmos. Cabe resaltar que este tópico es de vital importancia a la hora evaluar la adquisición e implementación de un sistema para la localización de fallas.

Se observa que los algoritmos en su totalidad utilizan como información de entrada las señales de Pre-Falla y Post-Falla en estado estable a la frecuencia fundamental. Además presentan múltiple estimación en la determinación de la ubicación de la falla, debido a que emplean el cálculo de la impedancia en cada sección de la línea desde la cabecera del circuito hasta el lugar probable de la falla incluyendo en el cálculo el valor de la resistencia de falla, generando incertidumbre acerca de la localización real. Por otra parte se observa que en su mayoría los algoritmos emplean los modelos de línea corta, presentando mayor simplicidad en los cálculos y la carga computacional pero comprometiendo la precisión del algoritmo. Finalmente, se destaca que la mayor parte de los algoritmos contemplan las principales características de los sistemas de distribución siendo este tópico de gran importancia para el trabajo.

Con el objetivo de realizar algunas pruebas complementarias, se seleccionó el algoritmo de Ratan Das, puesto que contempla todas las características de un sistema de distribución y presenta buena relación entre la simplicidad en la implementación y la precisión de los resultados. Otra razón a comentar es la disponibilidad de acceso a la información, pruebas y resultados relacionados con el algoritmo [Villamizar y Quiñones, 2005]. Estas pruebas se presentan en el capítulo 6.

## 2.6. Óptica estadística (Técnicas basadas en el conocimiento)

Para poder enfrentar apropiadamente las nuevas condiciones del sistema, es indispensable emplear técnicas modernas de análisis y herramientas complementarias para la detección, localización e identificación de las fallas en los sistemas de energía eléctrica, sin desvincularse totalmente de los avances tecnológicos en el campo del hardware, con el fin de garantizar la efectiva y útil implementación. Dichas técnicas permiten utilizar información adicional de la respuesta del sistema ante situaciones de falla [Lee et al., 2004].

El aprendizaje supervisado o clasificación es un problema de gran interés para los expertos en

Inteligencia Computacional. El diagnóstico de un evento dentro del sistema, se asocia rápidamente con la de detección de perturbaciones, errores y defectos. Cuando el diagnóstico se basa en observaciones múltiples<sup>3</sup>, se agrupa y discriminan los objetos para llegar a la construcción de las clases. Las observaciones son instancias de las clases que se desean modelar e incluyen una serie de atributos o características que representan su comportamiento. En consecuencia, la clasificación se puede ver como la agrupación de individuos que tienen características similares. El objetivo de la clasificación es describir, discriminar y asociar de forma precisa cada clase utilizando los atributos de los datos de entrada según sus similitudes [Isaza, 2004].

Visto así, la clasificación consiste en hacer un mapeo de individuos en el espacio *d-dimensional*<sup>4</sup>, al espacio *c-dimensional*<sup>5</sup> de clases, basándose en alguna medida de similitud. La forma en que se hace el análisis y agrupación de los mismos, depende del tipo de metodología empleada para la clasificación. En general los métodos de clasificación se basan en optimizar una función objetivo que representa el tipo de similitud que se analice en los objetos. Es así como hay técnicas que optimizan la distancia entre objetos del mismo grupo y otras que optimizan la pertenencia de los objetos a una clase.

El modelo de clasificación puede construirse a partir de la asesoría de especialistas en el tema o con la extracción manual del conocimiento a pesar de la dificultad que esto implica. No obstante, si se dispone de suficiente información registrada en una base de datos, el modelo de clasificación se puede construir generalizando a partir de ejemplos específicos mediante algún proceso inductivo. Las técnicas inductivas de clasificación se basan en el descubrimiento de patrones en los datos de entrada, por lo que se debe disponer de suficientes casos de entrenamiento para obtener un modelo de clasificación fiable. Se necesita gran volumen de datos para diferenciar entre patrones válidos y erróneos, por lo tanto, el modelo así obtenido puede servir para clasificar casos cuyas clases se desconozcan o simplemente, para comprender mejor la información de la que se dispone.

Los casos de entrenamiento utilizados en la construcción del modelo de clasificación suelen expresarse en términos de un conjunto finito de propiedades o atributos representados con valores discretos. Las categorías a las que han de asignarse los distintos casos se deben establecer de antemano. En general, estas clases deben ser independientes<sup>6</sup> y deberán ser discretas.

---

<sup>3</sup>Denominados también como *datos de entrada* o *datos de entrenamiento*.

<sup>4</sup>*d* corresponde al número de descriptores para el análisis.

<sup>5</sup>*c* corresponde al número de clases para el análisis.

<sup>6</sup>Algunas veces pueden establecerse jerarquías entre clases.

Trabajos relacionados con la localización de fallas y afines se han consultado durante el desarrollo de este trabajo. Se destacan investigaciones realizadas a partir de la aplicación de técnicas basadas en el conocimiento como LAMDA [Aguilar y Aguado, 1999], [Isaza, 2004], [Barrera, 2006] y Máquinas de Soporte Vectorial [Rodríguez, 2006], [Morales y Gómez, 2005], [Meléndez et al., 2005], además de trabajos realizados donde se aplican técnicas estadísticas para dar solución a la problemática planteada.

### 2.6.1. Técnicas Estadísticas (Reconocimiento de Patrones)

El planteamiento estadístico del problema de la clasificación<sup>7</sup> se explica de la siguiente manera. Se dispone de un conjunto amplio de observaciones de dos o más poblaciones diferentes. En este momento se realizan dos consideraciones, la primera, cuando se conoce la distribución para cada observación y su grado de pertenencia a las poblaciones contempladas. En este caso se admite una solución general. Sin embargo, la segunda consideración establece que los parámetros de las distribuciones son desconocidos y deben estimarse a partir de los datos. Para este caso, la dificultad en la discriminación se convierte en predecir nuevos valores de la variable dependiente a partir de un nuevo valor conocido de la variable independiente. Finalmente, se desea clasificar una nueva observación en una de las poblaciones.

El problema de la discriminación aparece en muchas situaciones en que se necesita clasificar elementos con información incompleta. En ingeniería el problema de la discriminación se ha estudiado con el nombre de reconocimiento de patrones<sup>8</sup>, con el fin de diseñar máquinas capaces de clasificar de manera automática [Peña, 2002]. Actualmente se considera las técnicas para el reconocimiento de patrones como una disciplina de la Inteligencia Artificial-IA. Para justificar este hecho se parte de dos principios:

- **Adquisición y representación del conocimiento:** Esta actividad se realiza a partir de la transferencia de información de una o más fuentes a un sistema informático para que sea procesado. Requiere patrones de una clase conocida para establecer un patrón prototipo de cada clase.
- **Aprendizaje:** Implica la adaptación a cambios para poder llevar a cabo una tarea. Es necesario plantear un algoritmo de aprendizaje para calcular el conjunto de patrones prototipo que caracterizan cada una de las clases a discriminar.

---

<sup>7</sup>Denominado también *discriminación*.

<sup>8</sup>Pattern Recognition en inglés.

El análisis de conglomerados<sup>9</sup> tiene como objeto congrega elementos en grupos en función de las similitudes entre ellos, de manera tal que los elementos correspondientes a un mismo grupo deben mostrar un alto grado de homogeneidad interna (dentro del conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados), siendo evidente gráficamente la cercanía de los elementos de un grupo y la lejanía entre elementos de diferentes grupos<sup>10</sup>, tal y como se muestra en la figura 2.2.

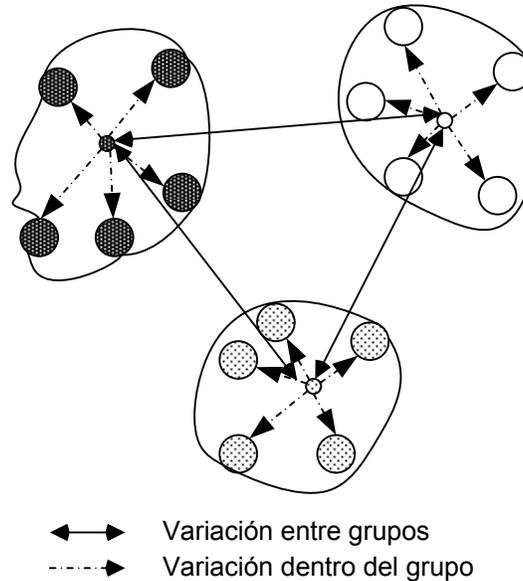


Figura 2.2: Diagrama de conglomerados que muestra la variación dentro y entre conglomerados.

Este análisis se conoce también como método de clasificación automática no supervisada<sup>11</sup> o reconocimiento de patrones sin supervisión, taxonomía numérica o técnica de cluster [Peña, 2002], [Diaz, 2002].

### Objetivos del reconocimiento de patrones

Las técnicas para el reconocimiento estadístico de patrones generalmente persiguen los siguientes objetivos:

1. La identificación de una estructura natural de los objetos, es decir, el desarrollo de una tipología o clasificación de objetos.
2. La búsqueda de esquemas conceptuales que expliquen el agrupamiento de algunos objetos.

<sup>9</sup>La palabra conglomerado es la traducción más cercana al término *cluster*.

<sup>10</sup>"Son regiones continuas de un espacio que contiene una densidad relativamente alta de puntos, las cuales están separadas por regiones que contienen una densidad relativamente baja de puntos "[Peña, 2002]. Everitt 1980.

<sup>11</sup>No supervisados se aplica para distinguirlos del análisis discriminante.

3. La formulación de hipótesis mediante la descripción y exploración de los grupos conformados.

Partiendo de que uno de los propósitos del reconocimiento de patrones es la comparación de objetos basados en algunas variables, éste análisis no está soportado sobre la base teórica que permita hacer inferencia y por ello suele calificarse como una técnica exploratoria. La solución no es única, las agrupaciones resultantes dependen de varios elementos en el procedimiento y de las variables utilizadas para su construcción. Conviene, sin embargo, estar siempre alerta ante el peligro de obtener, como resultado del análisis, no una clasificación de los datos sino una disección de los mismos, en distintos grupos que solo existen en la memoria del computador. *El conocimiento que el analista tenga acerca del problema decidirá cuales de los grupos obtenidos son significativos y cuales no.*

### Clasificación de los métodos para el reconocimiento de patrones

Existe una gran variedad de algoritmos establecidos para la conformación de conglomerados, y cada uno de estos métodos tiene una perspectiva diferente, produciendo resultados generalmente distintos al ser aplicados sobre el mismo conjunto de datos. Es importante tener presente que al momento de aplicar algún método de reconocimiento de patrones debe existir concordancia entre la medida de similaridad utilizada, las variables que se consideran y la naturaleza de la tipología esperada [Suárez y Salamanca, 2006].

- **Métodos Jerárquicos:** Se parte de la matriz de distancias o similaridades entre los elementos de la muestra. Se construye una jerarquía basada en estas distancias, la cual posee la característica de que en cada paso del algoritmo solo un objeto cambia de grupo, mientras que los grupos se encuentran establecidos en los pasos anteriores. El proceso de agrupamiento se hace de una manera acumulativa o por procesos de división [de Sá, 2001].
  - En el tipo acumulativo se comienza con tantas clases como elementos existan, con el fin de obtener sucesivamente los elementos más cercanos y formar con ellos una nueva clase. El algoritmo termina cuando todos los elementos queden ubicados en un sola clase.
  - Los métodos de división inician con todos los elementos dentro de una misma clase y en cada paso del algoritmo se divide la clase más heterogénea. El algoritmo termina cuando todos los elementos queden ubicados en clases lo suficientemente homogéneas.
- **Métodos de Partición** Difieren de los métodos jerárquicos debido a que existe un conocimiento a priori del número de grupos a conformar. Es un método muy utilizado cuando la cantidad de datos a clasificar es elevada. Adicionalmente se utiliza para refinar una clasificación realizada por un método jerárquico. Otra diferencia con las técnicas jerárquicas es que

la ubicación de un objeto dentro de un grupo no es definitiva. Esta técnica de clasificación presenta el siguiente algoritmo [Duda et al., 2000], [Rencher, 1995].

- Partición del conjunto de elementos en algún número específico de grupos. A cada uno de estos grupos se le calcula el centroide.
- Ubicación de cada elemento en el conglomerado cuyo centroide este más cercano.
- Cálculo de los nuevos centroides, luego de la comparación de todos los centroides con cada elemento.
- Se finaliza cuando todos los elementos se encuentran incluidos en un grupo de manera definitiva.

La estrategia planteada para resolver el problema de la localización fallas se basa en la caracterización de la respuesta del sistema en presencia de fallas, a partir de la extracción de atributos que describan de forma adecuada el comportamiento del sistema bajo estas condiciones. Inicialmente, se toman los datos y se aplica la técnica denominada análisis de componentes principales-PCA<sup>12</sup> con el fin de realizar un análisis exploratorio de datos-EDA<sup>13</sup> previo al reconocimiento de patrones. Posteriormente, se procede a realizar el análisis de la información extractada en los descriptores, estableciendo reglas y condiciones para la conformación de zonas características y relaciones de homogeneidad entre los grupos, esto se logra con la aplicación de técnicas para el análisis de conglomerados. El conjunto de datos recolectados puede ser útil debido a que permite establecer de manera probable el lugar de ocurrencia de una falla luego que ésta sucede [Suárez y Salamanca, 2006].

---

<sup>12</sup>Principal Component Analysis en inglés.

<sup>13</sup>Exploratory data analysis en inglés.



# Capítulo 3

## Análisis estadístico multivariado

### 3.1. Introducción

Este capítulo revisa los conceptos básicos y los algoritmos de la estadística multivariada utilizados en este trabajo.

La filosofía del análisis estadístico multivariado nace como respuesta a la necesidad de extraer conocimiento de información con alta dimensionalidad. Este hecho se pudo llevar a cabo a partir de los desarrollos realizados durante la revolución de la informática y ha facilitado tanto a científicos como a ingenieros almacenar y procesar grandes cantidades de datos. Para los propósitos de cualquier disciplina aplicada, la mayor parte de los fenómenos se analizan y modelan de forma multivariada con el fin de entender y caracterizar su comportamiento [Hair et al., 1999].

Los métodos estadísticos multivariados hacen posible plantear preguntas específicas y precisas de considerable complejidad en marcos idóneos, lo que posibilita llevar a cabo investigaciones significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural dentro del contexto en que normalmente ocurren.

### 3.2. Definición

El análisis multivariado se refiere a todos los métodos estadísticos que analizan simultáneamente medidas de cada individuo u objeto sometido a investigación; cualquier análisis simultáneo de dos o más variables<sup>1</sup> puede ser considerado como multivariado. El propósito del análisis multivariante

---

<sup>1</sup>Se denomina variable a cualquier atributo o característica observable en los elementos de una muestra o de una población.

es medir, explicar y predecir el grado de relación de los valores teóricos (combinaciones ponderadas de variables u observaciones) [Hair et al., 1999].

### 3.3. Conceptos básicos utilizados

#### 3.3.1. La matriz de datos

Las técnicas de análisis multivariado tratan con datos asociados a conjuntos de medidas sobre un número de individuos u objetos. El conjunto de individuos junto con sus variables se dispondrá en un arreglo matricial  $\mathbf{X}$  de dimensiones  $(n \times p)$ . Donde  $n$  corresponde al número de elementos y  $p$  corresponde al número de variables, tal y como se muestra en la ecuación 3.1 [Peña, 2002], [Díaz, 2002], [Rencher, 1995].

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (3.1)$$

Las variables pueden ser cuantitativas o cualitativas, y estas a su vez son continuas o binarias respectivamente. En el estudio solo se tendrán presente las variables cuantitativas y continuas. Cada una de estas  $p$  variables se denomina variable univariada y el conjunto de las  $p$  variables forman una variable multivariada [Peña, 2002].

#### 3.3.2. El valor teórico

Se refiere a una combinación lineal de variables con ponderaciones determinadas empíricamente. El investigador especifica las variables, mientras que las ponderaciones son objeto específico de determinación por parte de la técnica multivariada. El valor teórico capta el carácter multivariado del análisis, por tanto el valor teórico es el punto central del análisis, sin menospreciar la contribución de cada variable por separado.

#### 3.3.3. Vector de medias $\bar{\mathbf{x}}$

El vector de medias es para algunos estudios la medida de centralización más utilizada para la descripción de los datos multivariados (3.2). Es un vector de dimensiones  $(p \times 1)$ , cuyas componentes son las medias de cada una de las  $p$  variables utilizadas. El valor de cada componente es calculado con la ecuación (3.3).

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad (3.2)$$

$$\bar{x}_p = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.3)$$

### 3.3.4. Matriz de varianzas y covarianzas $\mathbf{S}$

La variabilidad de una variable respecto su media es cuantificada con la varianza y la dependencia lineal existente entre dos variables se mide con la covarianza. La matriz de varianzas y covarianzas (3.4) es una matriz cuadrada y simétrica de orden  $p$ . Los elementos de la diagonal principal corresponden a los valores de las varianzas, mientras que los elementos por fuera de ésta corresponden a los valores de covarianza entre las variables.

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (3.4)$$

La varianza de la  $i$ -ésima variable es calculada con la ecuación (3.5), donde  $\bar{x}_i$  es la media de la  $i$ -ésima variable. La covarianza entre la  $i$ -ésima y  $j$ -ésima variable es calculada con la ecuación (3.6), donde los subíndices  $i$  y  $j$  son utilizados para identificar las variables y  $k$  para las observaciones.

$$s_{ii} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \quad (3.5)$$

$$s_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad (3.6)$$

Otra ecuación para el cálculo de la matriz  $\mathbf{S}$  se presenta en (3.7), esta ecuación se encuentra en términos de las  $i$  observaciones (vectores).

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \quad (3.7)$$

### 3.3.5. Matriz de correlación $\mathbf{R}$

Para determinar la interdependencia lineal entre las variables se utiliza el coeficiente de correlación muestral  $r$ , el cual se debe analizar teniendo presente su magnitud y signo así: valores

cercanos a 1 o -1 indican que los puntos están a lo largo de una línea recta con pendiente positiva o negativa respectivamente mientras que valores cercanos a cero indica que no hay relación lineal, sin descartar que exista otro tipo de relación [Diaz, 2002]. La matriz de correlación es una matriz cuadrada y simétrica de orden  $p$ . Los elementos de la diagonal principal son iguales a 1, mientras que los elementos por fuera de ésta corresponden a los coeficientes de correlación lineal entre las variables. El arreglo matricial de los coeficientes de correlación se muestra en (3.8).

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (3.8)$$

El coeficiente de correlación muestral entre la  $i$ -ésima y  $j$ -ésima variable se calcula con la ecuación (3.9).

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \quad -1 \leq r_{ij} \leq 1 \quad (3.9)$$

En el caso en que las variables presenten diferentes escalas de medida o unidades, se puede incurrir en el error de descartar una de estas en la estructura de variabilidad por tener diferentes grados de dispersión. En particular, si se cambian las unidades, los resultados podrían cambiar completamente. Con el fin de evitar este tipo de inconvenientes y poder realizar comparaciones adecuadas, se sugiere aplicar algún tipo de transformación a las variables.

### 3.3.6. La distancia

El concepto de distancia entre puntos se utiliza de forma alternativa para estudiar la variabilidad de un grupo de observaciones, determinando la cercanía entre dos o más datos, con el objeto de rechazar o no una hipótesis estadística de acuerdo con la discrepancia de la muestra. Las propiedades de la función distancia son [Peña, 2002], [Diaz, 2002]:

- $d(\mathbf{A}, \mathbf{B}) \geq 0 \quad \forall (\mathbf{A}, \mathbf{B})$
- $d(\mathbf{A}, \mathbf{B}) = 0 \quad \Leftrightarrow \quad \mathbf{B} = \mathbf{A}$
- $d(\mathbf{A}, \mathbf{B}) = d(\mathbf{B}, \mathbf{A})$
- $d(\mathbf{A}, \mathbf{C}) \leq d(\mathbf{A}, \mathbf{B}) + d(\mathbf{B}, \mathbf{C})$

Donde  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  son individuos de un grupo de datos que tienen coordenadas  $(a_1, \dots, a_p)$ ,  $(b_1, \dots, b_p)$  y  $(c_1, \dots, c_p)$  pertenecientes a  $\mathfrak{R}^p$  respectivamente. La notación  $d(\mathbf{A}, \mathbf{B})$  corresponde a la

distancia entre los individuos **A** y **B**. Las métricas más utilizadas en la estadística multivariada son la distancia Euclidiana y la distancia de Mahalanobis. Estas distancias se describen a continuación.

### Distancia Euclídea

Se utiliza en muchos casos, pero presenta dificultades cuando las variables a tratar poseen diferentes unidades de medida. Este problema se evita dividiendo cada variable por un término que elimine el efecto de escala, conduciendo a una familia de distancias ponderadas. La ecuación (3.10) define de forma matricial la distancia entre dos puntos, donde  $x_{hj}$  y  $x_{ij}$  son el observaciones que pertenecen al espacio  $\mathfrak{R}^p$ ,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  y  $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$ , mientras que la ecuación (3.11) define de forma matricial la distancia entre un punto y su vector de medias.

$$d_{ij} = [(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)]^{\frac{1}{2}} \quad h, i = 1, \dots, n \quad (3.10)$$

$$d_i = [(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}})]^{\frac{1}{2}} \quad i = 1, \dots, n \quad (3.11)$$

### Distancia de Mahalanobis

Se emplea cuando las variables del estudio se encuentran en diferentes escalas de medición y están correladas. El valor que representa la distancia entre dos individuos no solo cambia de acuerdo con las unidades de medida empleadas sino por el grado de asociación que hay entre estas variables (3.12). La distancia de Mahalanobis se emplea frecuentemente para medir la distancia entre una observación multivariada y el centro del grupo de procedencia (3.13).

$$d_{ij} = [(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)]^{\frac{1}{2}} \quad h, i = 1, \dots, n \quad (3.12)$$

$$d_i = [(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})]^{\frac{1}{2}} \quad i = 1, \dots, n \quad (3.13)$$

Con las métricas se evalúan dos aspectos importantes, el primero, concerniente a la cercanía entre puntos de un mismo grupo (homogeneidad interna), esto indica que tan compacto está un conjunto de datos. El segundo, relacionado con la distancia entre los elementos de los diferentes grupos (heterogeneidad externa), para garantizar que no se genere traslape entre grupos, por el hecho de estar relativamente cerca.

## 3.4. Supuestos utilizados

El análisis estadístico multivariado requiere de la utilización de algunos supuestos implícitos a las técnicas estadísticas convencionales, con el objetivo de contrastar el valor teórico del modelo con el de las variables individuales, debido a que el valor teórico tanto del modelo como el de las variables debe cumplir con los mismos supuestos [Hair et al., 1999]. A continuación se mencionan algunos de los supuestos que usa el análisis multivariado.

### 3.4.1. Normalidad

El primer supuesto realizado y quizás uno de los más importantes, es el supuesto que se refiere a la correspondencia del perfil de la distribución de los datos con la distribución normal. Si la variación respecto a ésta es lo suficientemente alta, todas las pruebas estadísticas realizadas no son válidas. Los valores de la mayoría de las medidas asociadas con objetos se aglomeran simétricamente en torno a un valor central específico. La mayor parte de estas medidas se ubican dentro de alguna distancia determinada respecto al valor central, a la izquierda o a la derecha, las demás se presentan de manera cada vez más escasa en tanto que la distancia al valor central es grande.

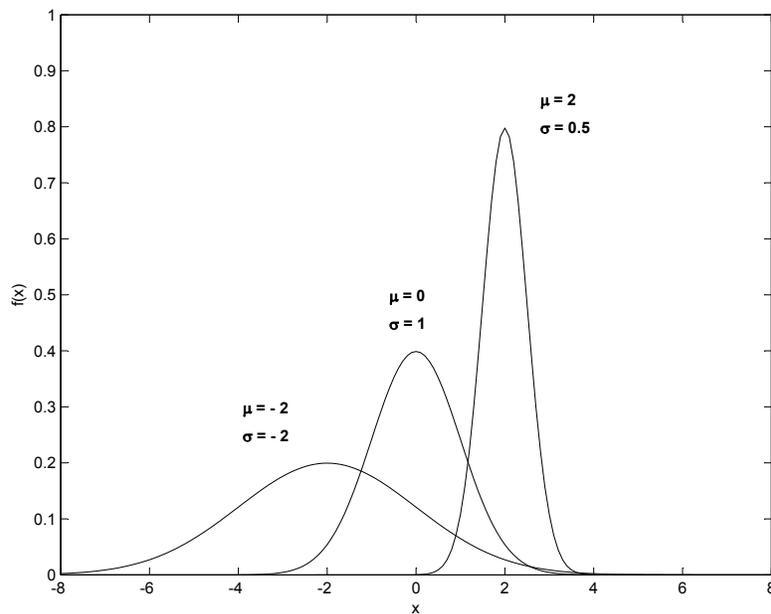


Figura 3.1: Funciones de densidad normal univariada.

La ecuación (3.14) representa la función densidad de probabilidad normal univariada, la cual es una de las distribuciones de probabilidad más utilizadas en estadística e ingeniería.

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\bar{x})^2}{2\sigma^2}\right\} \quad -\infty < x < \infty \quad -\infty < \bar{x} < \infty \quad \sigma^2 > 0 \quad (3.14)$$

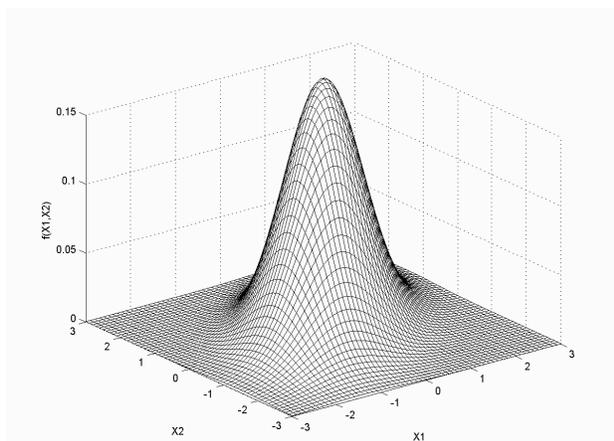
En la figura 3.1, se observan tres funciones de densidad normales con diferentes valores de medias y varianzas. Valores altos de varianza determina una gran dispersión de los datos alrededor de la media. En el caso contrario, la poca dispersión denota una gran tendencia de los datos hacia el valor definido por la media de la distribución. A menudo una muestra reúne una serie de elementos que aparentemente no poseen una relación entre sí.

### Normalidad multivariada

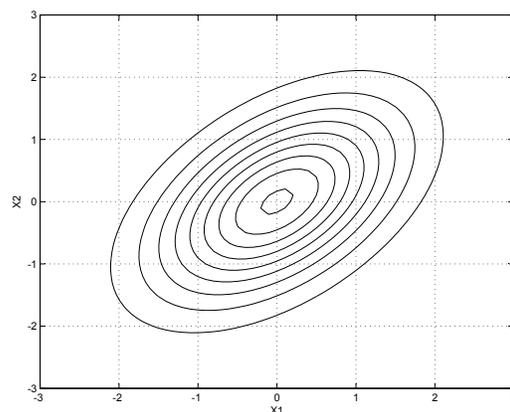
Es la combinación de dos o más variables que presentan distribución normal en sentido univariado y que sus combinaciones sean también normales. Por consiguiente si una variable es normal multivariada, es también normal univariada, sin ser necesariamente ésta suposición bidireccional. La ecuación (3.15) representa la función densidad de probabilidad normal multivariada con vector de medias  $\bar{\mathbf{x}}$  y matriz de varianzas y covarianzas  $\mathbf{S}$ .

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{S}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right\} \quad (3.15)$$

En la figura 3.2 se observa la gráfica de una distribución normal bivariada y su representación geométrica. En la figura 3.2a se realiza la representación tridimensional de la distribución, mientras que en la figura 3.2b se observa una familia de elipsoides concéntricos en  $\bar{\mathbf{x}}$ . El eje principal de cada elipsoide está ubicado sobre la línea que pasa a través de los puntos más distantes de la elipse.



(a)



(b)

Figura 3.2: Función de densidad normal multivariada. (a) Distribución normal bivariada. (b) Familia de curvas asociada.

### 3.4.2. Homocedasticidad

La homocedasticidad es un supuesto que trata las relaciones de dependencia entre variables. Dicho supuesto se refiere a la similitud de los niveles de variabilidad que exhiben las variables dependientes a lo largo del rango de su predictor. La homocedasticidad es deseable porque la varianza de la variable dependiente no debería concentrarse solo en un limitado rango de los valores independientes.

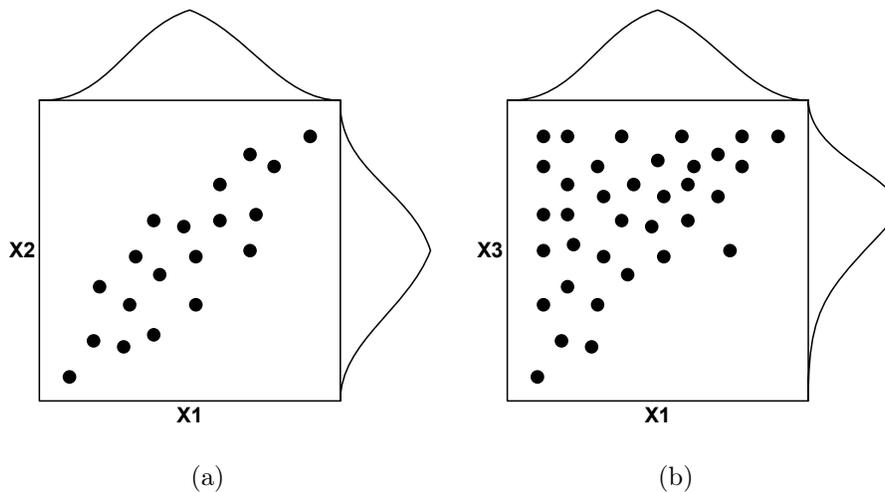


Figura 3.3: Gráfica de dispersión con relaciones de homocedasticidad y heterocedasticidad. (a) Homocedasticidad. (b) Heterocedasticidad

En la figura 3.3 se observa la dispersión de datos para tres variables ( $X1$ ,  $X2$ ,  $X3$ ) con distribución normal. Las variables  $X1$  y  $X2$  exhiben la misma dispersión de los datos reflejando una relación homocedástica tal y como se muestra en la figura 3.3a. Sin embargo existen casos donde las variables presentan diferentes niveles de dispersión, un ejemplo de este caso es el mostrado en la figura 3.3b, donde se observa la variación de la dispersión entre las variables  $X1$  y  $X3$  reflejando una relación heterocedástica, originada en la asimetría de la variable  $X3$ . El efecto de la heterocedasticidad es frecuentemente relacionado con el tamaño de la muestra, especialmente cuando se examina la dispersión de la varianza entre grupos.

### 3.4.3. Linealidad

La linealidad es un supuesto implícito de todas las técnicas multivariadas basadas en medidas de correlación, incluyendo el análisis de componentes principales. Los efectos no lineales no están representados en el valor de la correlación, por lo que es prudente revisar como impacta la correlación en los desplazamientos de la linealidad.

## 3.5. Planteamiento inicial

El análisis multivariable puede plantearse en dos niveles, el primero, consiste en la extracción de la información que contienen los datos disponibles (Análisis Exploratorio de Datos - EDA) y el segundo, en el cual se pretende generar conocimiento sobre el problema y obtener conclusiones sobre la población de los datos, a partir de la construcción de un modelo estadístico que permita explicar la forma como se generan y predicen los datos (inferencia) [Peña, 2002]. Tomando como base la propuesta para el análisis de datos multivariados mencionada anteriormente, se aplican tres técnicas para la construcción adecuada del modelo, la primera, Análisis de Componentes Principales - ACP, la segunda, k-medias y la tercera, Mezclas de distribuciones - MD. Con la aplicación de las dos primeras se realiza la exploración de los datos y la definición de las variables a utilizar, mientras que con la tercera se realiza la estimación de la función de probabilidad.

Como un paso adelante en el entendimiento de los datos multivariados se tiene que la exploración visual es una herramienta poderosa para el análisis de la información, esto ayuda al entendimiento de los mismos aunque el aumento del volumen y la complejidad de la información sean evidentes. La aplicación del ACP tiene como objetivo la estructuración de un conjunto de datos multivariable mediante la reducción del número de variables, a costa de una pérdida mínima de información [Wang et al., 2000]. La descomposición de los datos en clusters permite la conformación de grupos significativos de manera analítica, con el objetivo de clasificar los datos de una muestra según similitudes o afinidades. Adicionalmente el uso de modelos relativamente simples para cada estructura local, ofrece gran facilidad de interpretación, análisis y simplificación computacional [Wang et al., 1998].

## 3.6. Análisis de componentes principales - ACP

El ACP es un método estadístico que suministra información sobre la interdependencia lineal de las variables. Tiene como principal objetivo la estructuración de un conjunto de datos multivariados mediante la generación de nuevas variables que conservan la misma información y además que son incorrelacionadas entre si. Las nuevas variables son combinaciones lineales de las primeras y contienen la mayor parte de la variabilidad presente en el conjunto inicial [Diaz, 2002]. Las principales características acerca de la aplicación del ACP se resumen a continuación:

- Reducción de la dimensión del espacio donde están inscritos los datos.
- Transformación de las variables originales, en general correladas, en nuevas variables incorreladas, facilitando su interpretación.

- La variabilidad total de los datos no es alterada por transformaciones ortogonales.
- Determinación de relaciones poco evidentes entre las variables.

Las nuevas variables se denominan componentes principales - CP y los valores de las nuevas componentes se denominan puntajes de las componentes principales. Generalmente las dos primeras CP contienen la mayoría de la información, por lo tanto es común ubicar los datos en los dos nuevos ejes o componentes. A este plano se le conoce como *plano factorial*, y en el se pueden realizar algunas clasificaciones de los individuos, de las variables o de ambos y la detección de valores atípicos. Se recomienda el uso de ACP cuando la correlación entre variables es alta. Realizar un ACP sobre la matriz de correlación ( $\mathbf{R}$ ) de las variables originales, es equivalente a realizarlo con la matriz de varianzas y covarianzas ( $\mathbf{S}$ ) de las variables tipificadas.

### 3.6.1. Cálculo de las componentes principales

Como se explicó anteriormente y con el fin de evitar complicaciones generadas por malas interpretaciones debido al empleo de diferentes unidades, se sugiere trabajar con la matriz de correlación de las variables originales  $\mathbf{R}$  [Peña, 2002]. A continuación se presenta un algoritmo para la determinación de las componentes principales de un conjunto de variables que están fuertemente correlacionadas entre si. Cabe resaltar que este procedimiento aplica de la misma forma tanto para la matriz  $\mathbf{R}$  como para la matriz  $\mathbf{S}$ .

1. Calcular la matriz  $\mathbf{R}$  de las variables que están describiendo un contexto.
2. Calcular los valores propios de la matriz  $\mathbf{R}$  ( $|(\mathbf{R} - \lambda\mathbf{I})| = 0$ ).
3. Ordenar de forma descendente los valores propios.
4. Determinar los vectores propios asociados de la matriz  $\mathbf{R}$  ( $(\mathbf{R} - \lambda\mathbf{I})\mathbf{X} = 0$ ).

Donde,  $\mathbf{I}$  es la matriz identidad,  $\lambda$  son los valores propios de  $\mathbf{R}$  y  $\mathbf{X}$  es el vector propio asociado a cada valor propio. Cada CP es la combinación lineal de las variables originales cuyos coeficientes son los mismos del vector propio asociado.

### 3.6.2. Selección del número e interpretación de las CP

Es difícil encontrar un criterio estrictamente formal para determinar el número de CP a utilizar. El analista de los datos es quien debe determinar que cantidad de variación explicada se considera satisfactoria en el análisis y seleccionar el número de CP que cumplen con éste

requisito. Cada CP aporta una parte de variabilidad de los datos analizados, distribuyéndose en forma decreciente desde la primera hasta la última. La mejor forma de explicar el procedimiento anterior, es utilizando un diagrama de barras que muestra que cantidad de información contiene cada componente principal.

Un aspecto clave del ACP es la interpretación de los factores ya que no viene dada a priori, sino que será inducida por el analista tras observar la relación de los factores de las variables iniciales. Cabe resaltar que es de vital importancia el conocimiento que el experto tenga acerca de los datos materia de investigación.

Generalmente las primeras CP se consideran que poseen la mayor parte de la información de las variables originales, sin embargo las últimas componentes pueden poseer información importante [Rencher, 1995]. Debido a que los autovalores sirven para mostrar las varianzas de las componentes principales, las últimas componentes principales tienen varianzas más pequeñas. Si el valor de la componente es cercano a cero, la componente representa una relación lineal entre las variables que es esencialmente constante, por lo tanto, si el último autovalor está cerca de cero, indica la presencia de una colinealidad que puede suministrar una nueva información para el investigador.

### 3.7. Método de las K-medias

Se parte de que se quiere realizar una partición inicial de una muestra heterogénea de  $n$  individuos con  $p$  variables, para después modificar su configuración hasta obtener la fragmentación óptima en  $k$  grupos homogéneos, tal que cada elemento quede clasificado en uno sólo de los grupos. Este método es muy sensible a los valores iniciales dados, por lo que es conveniente utilizar los más adecuados. Una forma de aproximar los valores iniciales es mediante una clasificación obtenida por un algoritmo jerárquico. Para esto, una herramienta gráfica como el dendrograma sugiere un número de agrupaciones en cada paso, pero no define un número óptimo de grupos. Realmente, determinar esta cantidad depende de la aplicación y el criterio del analista de los datos. La estrategia propuesta para implementación del algoritmo *K-medias* es la siguiente:

1. Especificar el número de grupos ( $k$ ) u obtener una partición inicial de la muestra, puede aplicar un método jerárquico o el criterio del analista.
2. Determinar los centros de los  $k$  grupos, puede ser de forma aleatoria o a priori.
3. Tomar cada observación y calcular la distancia a cada centro de un grupo (3.11), (3.13).

4. Asignar cada observación al grupo cuya distancia al centro sea mínima.
5. Recalcular los centros de cada grupo.
6. Repetir los pasos 2, 3, 4 y 5 hasta que se satisfaga el criterio de parada (no ocurran más cambios en los grupos).

La idea es que la suma de los cuadrados de la distancia de todos los puntos que forman un grupo respecto a su centro sea mínima. A partir del agrupamiento realizado se inicia la identificación de las características predominantes en cada grupo, con el objeto de inferir y relacionar nuevos datos. El reconocimiento de patrones no tiene pretensiones inferenciales hacia una población a partir de una muestra, se emplea fundamentalmente como una técnica exploratoria (no presenta únicas soluciones).

### 3.8. Mezclas de distribuciones - MD

Las MD es una técnica de modelado estadístico con la cual se obtiene una estimación de la función densidad de probabilidad de los datos en una muestra aleatoria, la cual se representa como una suma ponderada finita de las componentes de densidad multivariadas. Esta técnica provee un sustento matemático para el modelado estadístico en una amplia variedad de fenómenos, reflejado en los últimos veinte años donde se han logrado avances considerables en el ajuste de modelos de mezclas [McLachlan y Peel, 2000]. Las MD se han aplicado con múltiples propósitos como: el modelado de la heterogeneidad de una población (Biología), manejo de datos faltantes, estimación de densidades de probabilidad y análisis de conglomerados (Estadística), reconocimiento de patrones (Imágenes) y obtención de consumos anómalos (Ingeniería Eléctrica) [Vargas y Cormane, 2004]. Dentro de las características principales de las MD se puede observar que no requieren un parámetro de suavizado de la función. Se puede establecer el número de términos de la mezcla y poseen un menor gasto computacional en comparación con otros métodos como la estimación de densidad Kernel y los histogramas [Martínez y Martínez, 2002]. La utilidad de la MD radica en el análisis y modelado de conglomerados (clusters). Al aplicar la técnica en el problema de localización de fallas, se utilizarán los conceptos sobre datos multivariadas. Las MD multivariadas se expresan como se muestra en la ecuación (3.16).

$$f(\mathbf{x}) = \sum_{g=1}^G p_g f_g(\mathbf{X}, \Theta) \quad (3.16)$$

Las MD sobre una muestra poblacional  $\mathbf{X}$  compuesta de  $n$  observaciones  $d$ -dimensionales, contiene  $G$  componentes  $f_g(\mathbf{X}, \Theta)$  referentes a la densidad de distribución multivariada seleccionada.

Cada componente de densidad  $f_g$  intenta describir el comportamiento de un grupo dentro de la población en el cual los datos relativos a dicho grupo poseen características similares, establecidas en el vector estimador  $\Theta$  correspondiente a los parámetros de cada distribución. Un estimador es un parámetro que define el comportamiento de los datos dentro de un grupo, y por tanto, su forma. Por ejemplo, los estimadores para una distribución de probabilidad normal son: el vector de medias  $\hat{\mathbf{x}}$  y la matriz de covarianza  $\hat{\mathbf{S}}$ , que definen el punto central de la distribución y la forma como se concentran los datos alrededor de dichos puntos respectivamente. La cantidad  $p_g$ , llamada proporción de la mezcla o coeficiente de mezclado, brinda información sobre la importancia del grupo dentro de la mezcla. Las condiciones que deben cumplir los coeficientes de mezclado están dadas en la ecuación (3.17).

$$\sum_{g=1}^G p_g = 1 \quad 0 < p_g < 1 \quad (3.17)$$

El propósito de la MD es identificar una cantidad desconocida de grupos en los cuales se aglomeren los datos de una muestra, es decir, la MD busca la homogeneidad dentro de una muestra inicialmente heterogénea. Para lograr este objetivo la MD utiliza el algoritmo Expectation-Maximization - EM para la estimación de parámetros. El algoritmo EM aparece como una forma de aplicación del método de la Máxima Verosimilitud - MV para la obtención de parámetros faltantes, permitiendo hallar de forma iterativa el vector estimador  $\Theta$  de MV para cada distribución  $f_g(\mathbf{X}, \Theta)$ , a partir del ingreso de valores iniciales para los parámetros de  $\hat{\Theta}^{(0)}$  [Peña, 2002]. El algoritmo trabaja con la función soporte  $L(\mathbf{Y}, \mathbf{Z}|\Theta)$  de la muestra, hallando la esperanza de las funciones de los datos ausentes  $\mathbf{Z}$  a partir del cálculo de la distribución de  $\mathbf{Z}$  con los valores iniciales  $\hat{\Theta}^{(0)}$  y los datos observados  $\mathbf{Y}$ .

$$L(\Theta|\mathbf{Y}) = E[l(\Theta|\mathbf{Y}, \mathbf{Z})] \quad (3.18)$$

Al resultado de esta operación se le denomina paso E (Estimación) mostrada en (3.18). En el paso M (Maximización) se maximiza la función  $L(\Theta|\mathbf{Y})$  respecto a  $\Theta$  con el fin de conseguir los estimadores MV a partir de la sustitución de las observaciones faltantes por estimaciones de sus valores. Sea  $\hat{\Theta}^{(i+1)}$  el estimador obtenido en el paso M, se retorna al paso E de manera iterativa hasta obtener la convergencia deseada [Figueiredo, 2002]. El método MV escoge como estimador de los parámetros aquel valor que hace máxima la probabilidad de que el modelo a estimar genere la muestra observada.

Es posible que dentro de una muestra exista la posibilidad de encontrar grupos representativos más pequeños que agrupen los elementos de la muestra, que permitan encontrar características

importantes que no necesariamente estén presentes en todas las observaciones disponibles. En esta instancia es aconsejable realizar un análisis de conglomerados o clusters que revelen la existencia de subconjuntos ocultos dentro de la muestra. Con la aplicación del algoritmo  $k$ -medias, se estiman los valores iniciales para los centros. El valor inicial de la matriz de covarianza se estima con la matriz identidad y el de los coeficientes de mezclado se calcula con la proporción de datos que poseen los grupos respecto a la muestra. Obtenidos los parámetros iniciales, se inicia la estimación de los parámetros con el algoritmo EM hasta alcanzar el valor de convergencia deseado. El resultado de este procedimiento son las estimaciones de los parámetros  $\widehat{\mathbf{x}}$ ,  $\widehat{\mathbf{S}}$  y  $p$  de cada grupo. Estos parámetros serán utilizados en la ecuación (3.16) con el fin de clasificar otras observaciones. A continuación se presentan los pasos del algoritmo EM.

1. Determinar el número de componentes de la mezcla ( $k$ -medias).
2. Determinar los valores iniciales de los parámetros de las componentes de densidad  $\widehat{\mathbf{x}}^{(0)}$ ,  $\widehat{\mathbf{S}}^{(0)}$  y  $p^{(0)}$ .
3. Calcular la probabilidad posterior para cada observación (Paso E), con las ecuaciones tal como se muestra en (3.19) y (3.20).

$$\widehat{\tau}_{ij} = \frac{\widehat{p}_i \Phi(\mathbf{X}_j | \widehat{\mathbf{x}}_i, \widehat{\mathbf{S}}_i)}{\widehat{f}(\mathbf{X}_j)} \quad (3.19)$$

$$\widehat{f}(\mathbf{X}_j) = \sum_{k=1}^C \widehat{p}_k \Phi(\mathbf{X}_j | \widehat{\mathbf{x}}_k, \widehat{\mathbf{S}}_k) \quad (3.20)$$

Donde  $\widehat{\tau}_{ij}$  representa la probabilidad posterior del punto  $\mathbf{X}_j$  correspondiente al  $i$ -ésimo término,  $\Phi(\mathbf{X}_j | \widehat{\mathbf{x}}_i, \widehat{\mathbf{S}}_i)$  es la densidad normal multivariable y  $\widehat{f}(\mathbf{X}_j)$  corresponde la MD estimada para el  $i$ -ésimo término evaluado en  $\mathbf{X}_j$ .

4. Actualizar  $\widehat{\mathbf{x}}$ ,  $\widehat{\mathbf{S}}$  y  $p$  de cada componente (Paso M), utilizando las ecuaciones (3.21), (3.22) y (3.23). Donde  $\widehat{\mathbf{x}}_i$ ,  $\widehat{\mathbf{S}}_i$  y  $p_i$  son las estimaciones actualizadas de los parámetros.

$$\widehat{p}_i = \frac{1}{n} \sum_{j=1}^n \widehat{\tau}_{ij} \quad (3.21)$$

$$\widehat{\mathbf{x}}_i = \frac{1}{n} \sum_{j=1}^n \frac{\widehat{\tau}_{ij} \mathbf{X}_j}{\widehat{p}_i} \quad (3.22)$$

$$\widehat{\mathbf{S}}_i = \frac{1}{n} \sum_{j=1}^n \frac{\widehat{\tau}_{ij} (\mathbf{X}_j - \widehat{\mathbf{x}}_i) (\mathbf{X}_j - \widehat{\mathbf{x}}_i)'}{\widehat{p}_i} \quad (3.23)$$

5. Repetir los pasos 3 y 4 hasta obtener la convergencia deseada.

Sea  $\widehat{\Theta}^{(i+1)}$  el estimador obtenido en el paso M, se retorna al paso E de manera iterativa hasta obtener la convergencia deseada, se decir, hasta que  $\left\| \widehat{\Theta}^{(i+1)} - \widehat{\Theta}^{(i)} \right\|$  sea lo suficientemente pequeño.

Posteriormente se realiza la organización de los grupos en clases basadas en la probabilidad de aparición de los datos en cada grupo, dada por el modelo de la mezcla (3.24).

$$\widehat{f}(\mathbf{X}) = \sum_{i=1}^C \widehat{p}_i \Phi_i \left( \mathbf{X} | \widehat{\mathbf{x}}_i, \widehat{\mathbf{S}}_i \right) \quad (3.24)$$

El uso de los algoritmos anteriores permite obtener estimadores que describan adecuadamente los datos de la muestra.



# Capítulo 4

## Metodología para la construcción del modelo estadístico

### 4.1. Introducción

Este capítulo describe de forma detallada los pasos empleados en la construcción del modelo estadístico, destacando los criterios utilizados.

Una forma de resolver el problema de la localización es conocer el comportamiento del sistema cuando se encuentra sometido a diferentes perturbaciones. Para lograr este propósito se propone realizar la caracterización de la respuesta del sistema en presencia de fallas, la cual, se ve reflejada en las señales de tensión y corriente. En consecuencia, la obtención de las señales es un paso fundamental en el momento extraer las características que describan el comportamiento del sistema. A continuación se efectuará una breve descripción por etapas de la metodología propuesta.

En la primera etapa, se selecciona el sistema prototipo. En la segunda etapa, se implementa el sistema prototipo en el software ATP/EMTP con el objeto de realizar simulaciones del sistema bajo condiciones de falla, conformando un banco de registros con las señales de tensión y corriente, obtenidas en cada simulación. Adicionalmente, se procede a realizar la adecuación de las señales de tensión y corriente, se obtiene el valor eficaz de las señales en el tiempo y se convierten al sistema en por unidad. En la tercera etapa, se plantea la caracterización del sistema de dos formas: la primera, de tipo determinista basada en el cálculo de descriptores monofásicos de la señal de tensión. La segunda, de tipo estadística la cual está basada en la aplicación de un análisis exploratorio de los datos, obteniendo finalmente descriptores característicos del comportamiento del sistema. En la cuarta etapa, se procede a realizar el análisis de la información extractada en los descrip-

tores, estableciendo relaciones, criterios y condiciones (homogeneidad y heterogeneidad) entre los grupos, conformando grupos característicos que relacionan zonas del circuito con la ubicación de la falla. Finalmente en la quinta etapa, se construye el modelo de mezclas teniendo presente los resultados obtenidos anteriormente. Adicionalmente, se realiza la validación del modelo a partir de la clasificación de los datos de falla (localización probable). Se realizan comparativos donde se cuantifican los resultados del modelo.

## 4.2. Sistema prototipo

Para la selección del modelo del sistema de distribución se tomó como criterio, que el circuito debía poseer las características descritas en el capítulo 2. El circuito seleccionado para la implementación de la metodología propuesta es el utilizado en [Das, 1998], para las pruebas de diferentes métodos de localización de fallas. El diagrama unifilar se muestra en la figura 4.1. En el apéndice A se presentan las especificaciones del sistema y los parámetros de las líneas y cargas, tales como: impedancias de secuencia, carga instalada, tipo de carga y factor de potencia.

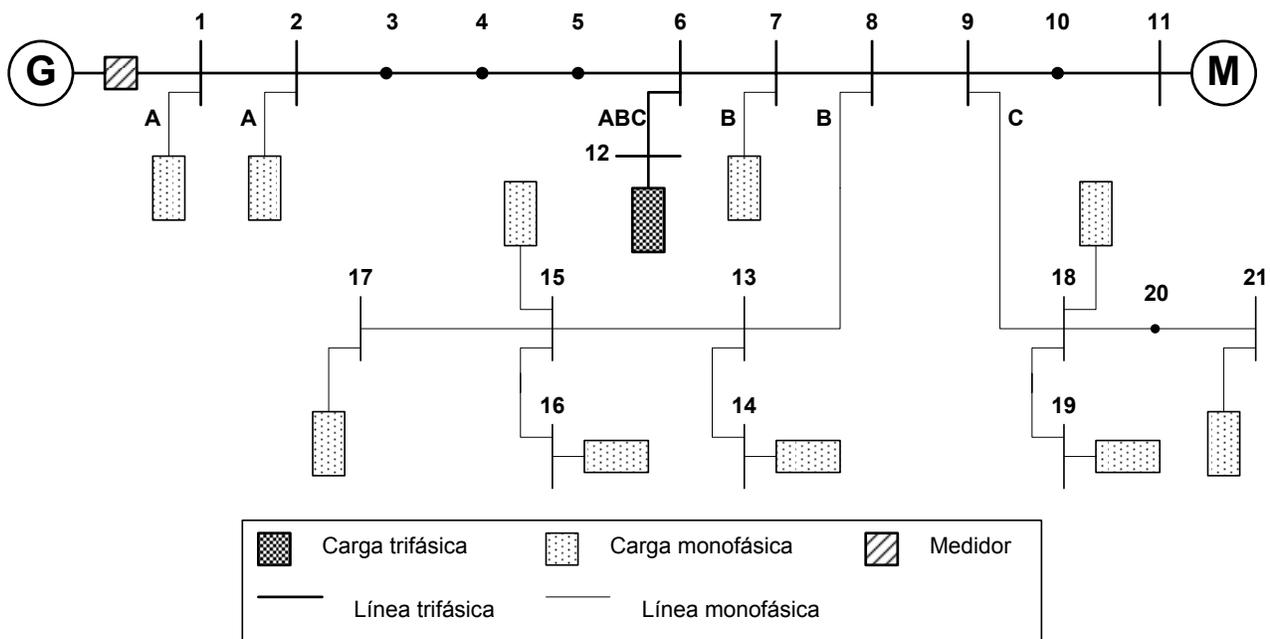


Figura 4.1: Diagrama unifilar del sistema de distribución.

El sistema está compuesto por ramales monofásicos y trifásicos, los cuales presentan diferentes secciones de conductor en varios tramos de las líneas. Se encuentra dividido en 21 nodos, 9 monofásicos y 12 trifásicos. Posee una tensión de operación de 25kV, 2 cargas monofásicas y 10 trifásicas distribuidas a lo largo del circuito. El equipo de medida se encuentra ubicado en la cabecera del circuito (Nodo 1).

### 4.3. Simulación del sistema

La simulación del sistema bajo condiciones de falla fue realizada con el software ATP/EMTP. Se construyó un banco de registros que contienen las ondas de tensión y corriente medidas en la cabecera del circuito obtenidas mediante simulación. Adicionalmente, se almacenaron las señales de tensión y corriente en valores RMS. Toda la información se almacena en archivos planos con el fin de facilitar el procesamiento y manejo de la misma. En la figura 4.2 se observa el diagrama esquemático del circuito sometido a simulación.

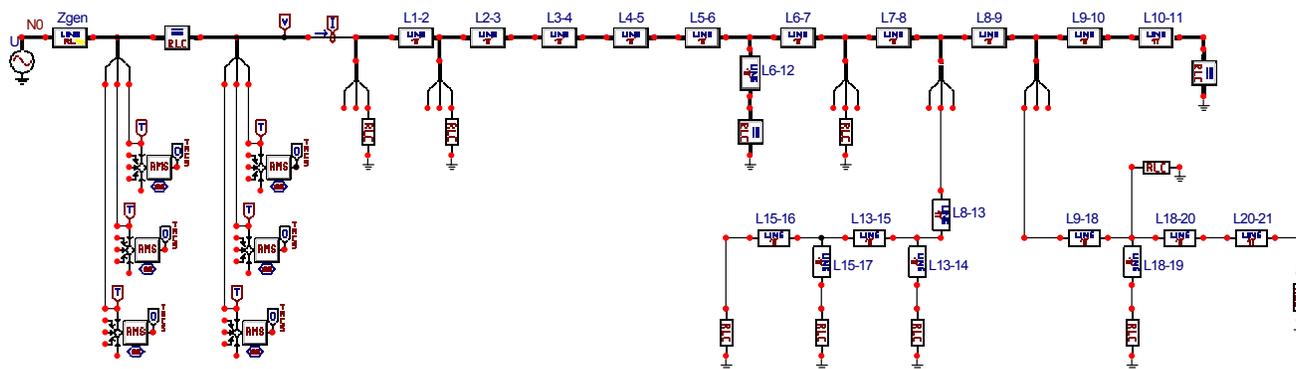


Figura 4.2: Sistema de distribución implementado en ATP/EMTP.

El procedimiento seguido para la obtención de la información consistió en realizar la simulación para los diferentes tipos de fallas mencionados en el capítulo 2, con valores de resistencia de falla entre  $0,05\Omega$  y  $50\Omega$ . Esta actividad se efectuó en cada uno de los 21 nodos del circuito.

Tabla 4.1: Simulaciones del sistema de distribución en ATP/EMTP

Tipo de Falla	Número de Simulaciones	Resistencia de Falla [ $\Omega$ ]
Monofásica	495	
Bifásica	396	
Bifásica-T	396	0,05 5 10 15 20 25 30
Trifásica	132	35 40 45 50
Trifásica-T	132	
<b>TOTAL</b>	<b>1551</b>	

En la tabla 4.1 se resumen los detalles de las simulaciones realizadas [Dagenhart, 2000] y se presentan los resultados de la simulaciones efectuadas. Se puede observar la totalidad de simulaciones realizadas por cada tipo de falla y valor de resistencia de falla. Es de notar la existencia de un número menor de simulaciones para fallas bifásicas y trifásicas. Este hecho se

debe a la presencia de ramales monofásicos en el sistema de distribución.

Como una alternativa para la extracción de información adicional de las formas de onda, se hace necesaria la implementación de una estrategia consistente en la adecuación de las señales de tensión y corriente, utilizando el algoritmo para el cálculo del valor eficaz de la señal en el tiempo. Lo anterior se puede observar como una caracterización aproximada de la señal a frecuencia fundamental en presencia de transitorios y distorsiones de onda [Arrillaga, 2000]. Finalmente, las señales procesadas se convierten al sistema en por unidad. La ecuación (4.1) se utiliza para obtener el valor eficaz de las señales. Donde  $N$  corresponde al número de muestras por ciclo y  $x_i$  son las muestras de la señal en el dominio del tiempo. Los valores eficaces se calcularon sobre una ventana de 128 muestras.

$$\mathbf{x}_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (4.1)$$

En la figura 4.3a se observa la forma onda de la señal de tensión obtenida luego de la simulación de una falla en una de las barras del sistema. Adicionalmente, en la figura 4.3b se muestra la misma señal en el tiempo pero esta vez en valores eficaces, en por unidad con un valor de tensión y potencia base de 25kV y 100MVA respectivamente.

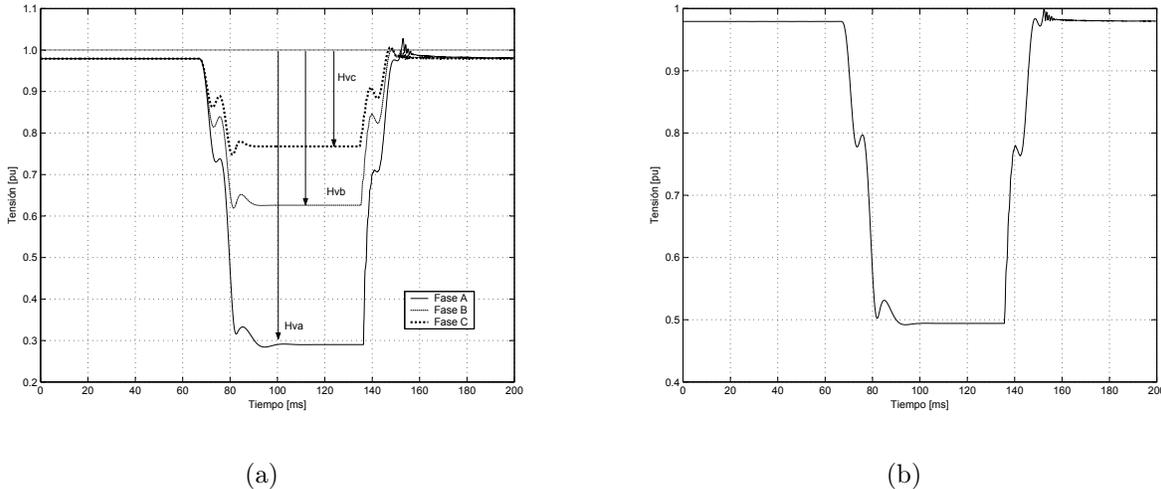


Figura 4.3: Señal monofásica de tensión durante una falla. (a) Forma de onda. (b) Valor eficaz.

Se puede observar que el valor eficaz de la tensión no se reduce inmediatamente pero toma un ciclo para la transición, por otra parte se observa que el valor eficaz no es constante durante la falla y no recupera su valor nominal de forma inmediata cuando usualmente la falla se despeja. Es de notar que la gráfica muestra las condiciones de prefalla, falla y postfalla de la señal.

## 4.4. Caracterización del sistema

Caracterizar el comportamiento de una situación o fenómeno real requiere de un número elevado de variables y relaciones simultaneas. Los descriptores son elementos que permiten cuantificar o cualificar las diferentes situaciones. A partir de su definición se puede proceder a clasificar e identificar las características presentes en cada aplicación. La naturaleza de los descriptores puede tener diferente origen y depende de los propósitos de la clasificación.

Los métodos utilizados para la localización de fallas en sistemas de distribución, que emplean medidas en un terminal de línea, se clasifican en dos grupos: los *métodos basados en el conocimiento - MBC* y los *métodos basados en el modelo - MBM* [Mora y Colomer, 2003]. Los primeros, contienen una base de conocimientos con información proveniente del procesamiento de los registros de tensión y corriente medidos en la subestación, mientras que los segundos utilizan el valor de la componente fundamental de las señales de tensión y corriente medidas en la subestación y el modelo del sistema de potencia, para calcular la impedancia aparente vista desde la subestación hasta el sitio de falla [Das, 1998]. Actualmente, los MBM son los más documentados para los sistemas de distribución, debido a que su implementación y operación no es tan costosa como los MBC. Sin embargo, la alta dependencia de un buen modelo y la múltiple estimación de la posible falla son sus grandes desventajas [Zhu et al., 1997], [Novosel et al., 1998]. Así mismo, las desventajas de los MBC es que suelen tener baja precisión y muy costosos en comparación con los MBM, debido al volumen de información que utilizan [Mora y Colomer, 2003].

En la mayoría de las empresas distribuidoras, solo se tienen señales de tensión y de corriente medidas en la subestación. Por esta razón, es importante desarrollar metodologías que permitan extraer información relevante a partir de estas mediciones. La caracterización permite aprovechar toda la información disponible en la subestación, sin aumentar significativamente el costo de la implementación de un método basado en el conocimiento. Por este motivo es útil determinar descriptores con los cuales se pueda realizar un adecuado modelado de la información y posteriormente generar conocimiento acerca del problema.

En este trabajo de investigación solo se utilizaron descriptores en estado estable para la construcción del modelo y se realizó un breve resumen acerca de trabajos realizados con descriptores en estado transitorio. Este hecho se justifica en parte en la necesidad de obtener una herramienta de fácil manejo y bajo costo, por lo tanto, la utilización del estado transitorio de la señal implica mayor costo en equipos de adquisición y procesamiento de datos, además de un más detallado conocimiento y análisis de los datos (*Complejidad vs. Nivel de detalle*).

#### 4.4.1. Descriptores en estado estable

Un descriptor es un atributo que permite la caracterización del comportamiento en el sistema. A continuación se describirán los descriptores propuestos para la caracterización de fallas, basados en dos tipos de análisis: el primero, basado en el análisis de la señal de tensión y corriente en estado estable antes, durante y después de la falla (determinísticos) y el segundo basado en la aplicación de técnicas para la búsqueda de relaciones entre datos (estadísticos).

Al momento de presentarse una falla en el sistema, se produce un incremento en la corriente y una reducción en los valores de tensión. Esto como respuesta del mismo a la perturbación. La reducción de la tensión se conoce como caída de tensión<sup>1</sup>. Según estándar internacional una caída de tensión se define como la reducción del valor eficaz de la tensión entre el 0,9 y 0,1 p.u. de su valor nominal, en un periodo comprendido entre 8,33 ms y 1 minuto, a frecuencia industrial [IEEEstd1159-1995, 1995]. Con la caracterización de la señal de tensión medida en la subestación, es posible describir una falla y encontrar su probable localización. Los descriptores planteados en este trabajo son una modificación de la propuesta encontrada en [Mora y Colomer, 2003]. En la figura 4.4 se realiza la representación gráfica de los descriptores utilizados para la construcción del modelo.

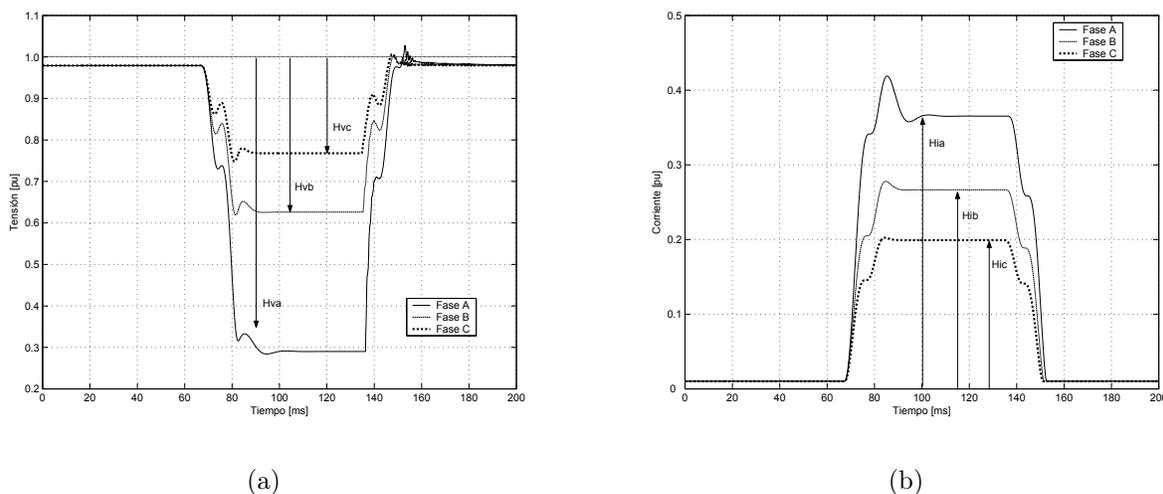


Figura 4.4: Descriptores monofásicos. (a) Descriptores de tensión. (b) Descriptores de corriente.

La magnitud de la caída de tensión es la medida en estado de falla, de la máxima reducción del valor eficaz de la señal de tensión por fase del sistema tal como se muestra en la figura 4.4a. Los descriptores de tensión se calculan con la ecuación (4.2). De forma similar, en la figura 4.4b se

<sup>1</sup>sag en inglés.

presentan los descriptores de corriente. La magnitud de la elevación de la corriente es la medida del valor eficaz de la señal de corriente por fase del sistema y se calcula con la ecuación (4.3). Los descriptores mencionados anteriormente se obtienen de forma sencilla, solo son necesarios los valores eficaces de las señales registradas en los medidores y un algoritmo para el procesamiento.

$$H_{v\phi} = 1 - \text{mín}(V_{rms\phi}) \quad (4.2)$$

$$H_{i\phi} = 1 - \text{máx}(I_{rms\phi}) \quad (4.3)$$

Para la obtención de los descriptores es importante tener presente los siguientes aspectos:

- *Tiempo de inicio de la caída de tensión* El inicio de este fenómeno se presenta a partir de que el valor eficaz de la tensión se reduce a un valor menor de 0.9 [p.u]. En ese instante se comienza a ubicar el estado estable de falla donde se presenta el mínimo valor de tensión y el máximo valor de corriente en las fases.
- *Ubicar el estado estable de la caída de tensión* Esto es necesario debido a que la parte transitoria del hueco de tensión varía con respecto al ángulo de incidencia de la falla.

La tabla 4.2 resume los descriptores de tensión y corriente utilizados en el modelo [Mora y Colomer, 2003], [López, 2006].

Tabla 4.2: Descriptores de tensión y corriente

Descriptor	Descripción	Unidades
$H_{va}, H_{vb}, H_{vc}$	Magnitud de la caída de tensión por fase	p.u
$H_{ia}, H_{ib}, H_{ic}$	Magnitud de la elevación de la corriente por fase	p.u

#### 4.4.2. Descriptores en estado transitorio

Las señales de tensión y corriente medidas en la subestación, presentan fuertes oscilaciones en el momento en que ocurre una perturbación en la red, estas oscilaciones se conocen como transitorios. El análisis de los transitorios se realiza con el objeto de extraer las diferentes componentes de frecuencia de las señales del sistema y asociarlas a una distancia eléctrica de posible ocurrencia de la falla.

El comportamiento del sistema en estado transitorio se puede modelar y analizar si se conoce el tiempo de amortiguamiento y la frecuencia de oscilación de la señal transitoria.

Estas características se pueden tomar como descriptores para facilitar la comprensión del tema. Los descriptores se encuentran directamente relacionados con la distancia a la falla. En los sistemas de potencia existe una relación entre los parámetros resistivos ( $R$ ), inductivos ( $L$ ), capacitivos ( $C$ ), la frecuencia de oscilación ( $f$ ) y la distancia. Ese hecho obedece a las características eléctricas y magnéticas inherentes a los elementos que constituyen el sistema [Mora et al., 2005].

La relación entre los parámetros  $L$  y  $C$  definen la frecuencia natural del transitorio, por esta razón este descriptor permite encontrar la distancia a la falla sin considerar los efectos resistivos. Mientras que para determinar la duración de los eventos transitorios se debe tener presente los parámetros  $R$ ,  $L$  y  $C$  del circuito en términos de sus constantes de tiempo [Greenwood, 1991]. Para obtener las componentes de frecuencia presentes en los transitorios, se utilizan dos herramientas para el análisis de señales: la transformada wavelet (TW) y la transformada rápida de Fourier (FFT).

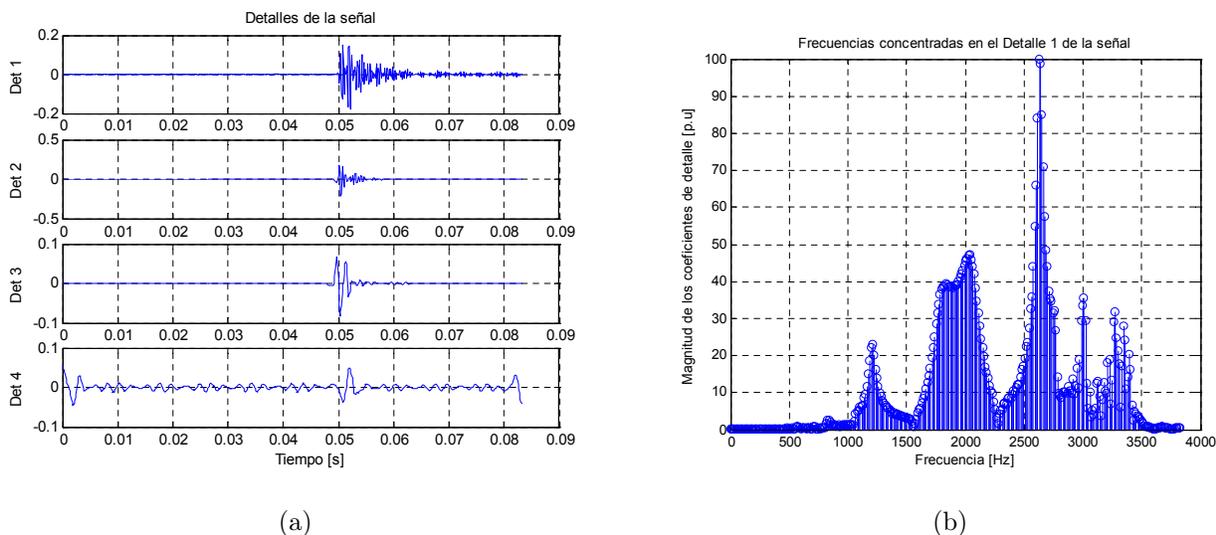


Figura 4.5: Tratamiento de la señal transitoria. (a) Coeficientes de detalle (TW). (b) Componentes de frecuencia (FFT).

### Transformada Wavelet (TW)

Este método de análisis de señales, filtra una señal en el dominio del tiempo mediante filtros que eliminan componentes de alta y baja frecuencia de la señal. El filtrado se aplica de manera secuencial, para las señales resultantes del proceso de filtrado anterior. Este procedimiento continúa hasta que la señal se ha descompuesto en un cierto número de niveles de detalle predefinidos, tal y como se muestra en la figura 4.5a. Como resultado del proceso de descomposición, se obtiene un conjunto de señales en diferentes bandas de frecuencia que representan a la señal original [Vega

et al., 2005]. A través de la TW se logra una buena representación tanto en tiempo como en frecuencia de la señal, ya que se puede determinar el intervalo de tiempo en el cual aparecen determinadas componentes espectrales.

### Transformada Rápida de Fourier (FFT)

Los detalles obtenidos con la TW se analizan con la transformada de Fourier (figura 4.5b), ya que ésta permite analizar señales no estacionarias dividiendo la señal original en pequeñas partes que se asumen estacionarias [Oppenheim y Willsky, 1998]. La principal ventaja de la FFT frente a la TW es la exactitud que ofrece esta para ubicación en el dominio de la frecuencia. En la tabla 4.3 se muestran las características que presentan cada una de las transformadas.

Tabla 4.3: Características de la TW y la FFT

<b>TW</b>	<b>FFT</b>
Multiresolución en tiempo y frecuencia de señales no estacionarias	Multiresolución en tiempo y frecuencia de señales no estacionarias con ventanas fijas
El dominio tiempo-frecuencia permite tener información de la frecuencia en cada instante	Localización de cambios abruptos de una señal en el dominio del tiempo
Resolución de tiempo y frecuencia variable	Baja resolución para señales de alta frecuencia

El procedimiento para la extracción de la frecuencia y el tiempo de amortiguamiento de la señal transitoria se describe a continuación y se resume en el diagrama de flujo de la figura 4.6:

1. Medición de señales de tensión. Normalmente, las señales de corriente no permiten observar el fenómeno transitorio con suficiente claridad.
2. Aplicación de la TW. Se obtiene un nivel de descomposición y un número de detalles deseado.
3. Determinación del detalle con mayor índice Wavelet. Este detalle se utiliza para determinar el tiempo de duración y la frecuencia del transitorio producido por la falla del sistema.
4. Aplicación de la FFT. Calculo de las componentes de frecuencia, del detalle que posee el mayor índice Wavelet.
5. Cálculo de la duración del transitorio. Definido como el tiempo en que el valor pico de la señal transitoria cae debajo del 2% de su valor.

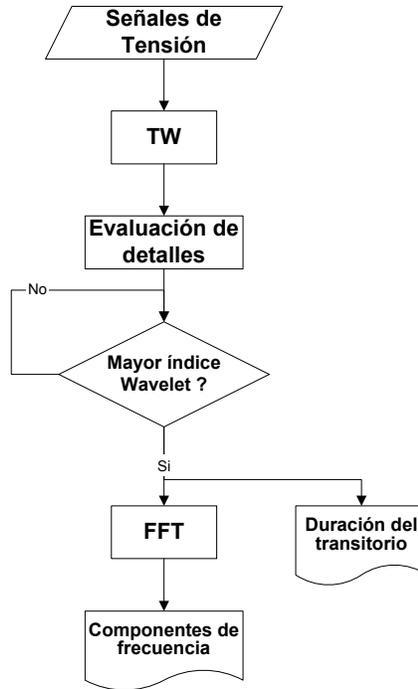


Figura 4.6: Procedimiento para la extracción de descriptores de estado transitorio.

## 4.5. Análisis de los datos

Esta sección tiene como objetivo principal, realizar el análisis exploratorio de datos con el fin de reconocer patrones de comportamiento característicos estadísticamente soportados, estableciendo relaciones, criterios y condiciones entre los datos que permitan la construcción del modelo de mezclas. A partir de los resultados arrojados por el EDA se indagará acerca de la información contenida en los datos facilitando la comprensión y el análisis de los mismos.

Inicialmente se plantea el análisis de correlación multivariado para determinar si se hace necesario la utilización del ACP en los datos. El grupo de variables empleadas describe el comportamiento de la falla monofásica en el sistema. Este procedimiento se muestra en la figura 4.7a. Se observa claramente la baja interdependencia lineal entre las variables (baja correlación), puesto que los valores fuera de la diagonal principal son mucho menores que  $\|1\|$ . El cálculo de los valores propios permite construir el diagrama mostrado en la figura 4.7b con el cual se puede determinar el porcentaje de información que explica cada componente. Se observa que la información se encuentra contenida de forma proporcional en dos (alta explicación) de las tres componentes. Esta proporción equivale al 78,75% del total de la información. En consecuencia una reducción de la dimensionalidad de las variables (excluir la última componente) implicaría una pérdida aproximada del 21,25% de la información. En estos casos la pérdida de información es considerable y coloca en tela de

juicio los resultados del procedimiento, por lo que es innecesario reducir la dimensionalidad del modelo. Para este caso la visualización del plano factorial con las tres componentes equivale a la visualización en tres dimensiones de las variables originales. Por lo tanto, se sugiere utilizar directamente los tres descriptores de tensión como variables de entrada linealmente independientes. Este resultado no justifica la utilización del ACP para la reducción de la dimensionalidad del modelo.

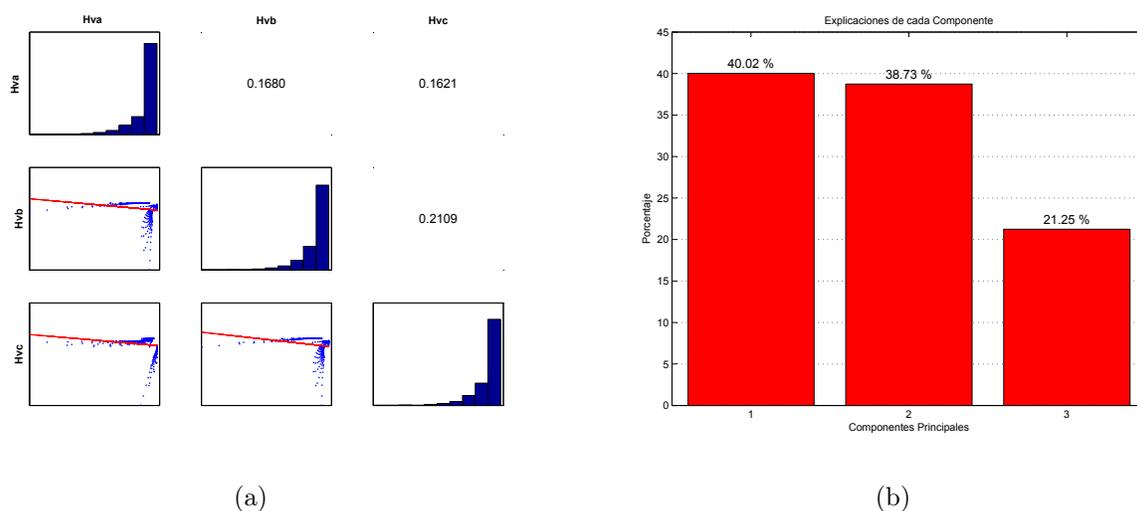


Figura 4.7: Resultados del ACP en los descriptores para fallas Monofásicas. (a) Análisis de correlación multivariado. (b) Porcentaje de información en cada CP.

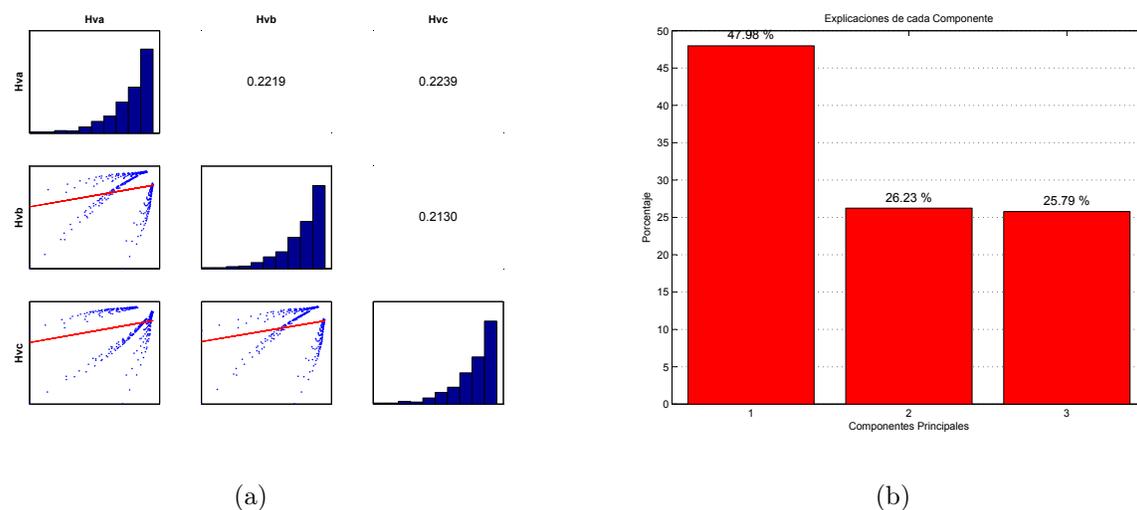


Figura 4.8: Resultados del ACP en los descriptores para fallas Bifásicas a tierra. (a) Análisis de correlación multivariado. (b) Porcentaje de información en cada CP.

Otra forma de notar el resultado anterior es analizando las figuras 4.8a y 4.8b, donde se observa que para el tipo de falla bifásica a tierra, la información se encuentra contenida de forma similar a

la falla monofásica. Para este caso la pérdida de información equivale al 25.79%, si se excluye la última componente. Este hecho refleja la independencia intrínseca que presentan los datos, la cual fue corroborada con el análisis de correlación y la aplicación del ACP.

## 4.6. Construcción del modelo de mezclas

El objetivo de esta sección es definir los grupos de datos característicos asociados a las zonas de falla y posteriormente utilizarlos en la construcción del modelo. Dicha agrupación debe realizarse con los datos del grupo de descriptores que mejor se ajusten a los datos obtenidos en las etapas anteriores. Las mezclas de distribuciones-MD, es una técnica que permite la comparación entre las características de un dato y los rasgos particulares de cada grupo presente, determinando la probabilidad de pertenencia al mismo.

### 4.6.1. Estructura del modelo

El modelo propuesto utiliza tres etapas diferentes de clasificación - EC, la primera denominada EC1, con la cual se identifica la fase (caso monofásico) o fases (caso bifásico) en la que sucede la falla. La segunda denominada EC2, en esta etapa se identifican los grupos asociados a cada uno de los rangos de valores de resistencia de falla. Por último la EC3, con la cual se identifica la zona de ocurrencia del evento dentro del sistema. Como resultado de la construcción del modelo se obtienen los valores de los parámetros  $\mathbf{p}$ ,  $\bar{\mathbf{x}}$ ,  $\mathbf{S}$  que definen el modelo de mezcla para cada uno de las etapas. Cabe anotar que las etapas EC1, EC2 y EC3 aplican para los tipos de falla monofásica y bifásica, mientras que para las trifásicas únicamente aplican las etapas EC2 y EC3. El procedimiento descrito anteriormente se puede observar en el diagrama de flujo mostrado en la figura 4.9.

#### ***EC1-Clasificación de los datos según la fase fallada***

Los datos de entrenamiento se utilizan para estimar los grupos que mejor representan el comportamiento de las fases en el momento de una falla. El proceso inicia con la estimación de los centros de los grupos utilizando el método *k-medias* para el análisis de conglomerados. Seguidamente se definen los valores iniciales para las matrices de varianzas y covarianzas de cada grupo, en este caso se utiliza la matriz identidad. Por último, se toman de forma *a priori* los valores iniciales para los coeficientes de mezclado, debido al desconocimiento de los mismos. Es decir, no se tiene información acerca de los parámetros de los grupos ni de la relación de estos con las características de la respuesta del sistema (frecuencia de ocurrencia de los diferentes tipos de falla, valores de resistencia y la ubicación del evento para este circuito). El procedimiento mostrado anteriormente

se emplea en las siguientes etapas de forma similar.

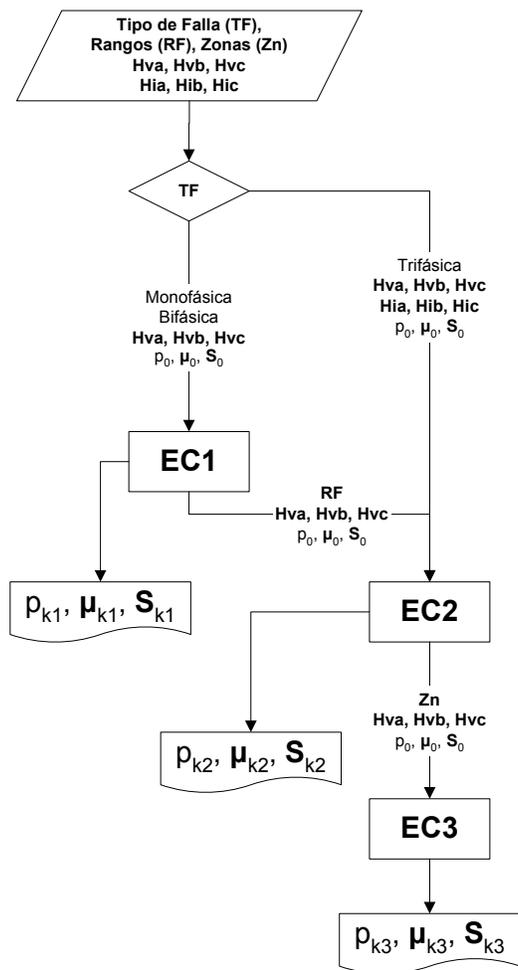


Figura 4.9: Diagrama de flujo de la construcción y entrenamiento del modelo estadístico.

### ***EC2-Clasificación de los datos según el valor de resistencia de falla asociada***

La información acerca del valor de resistencia de falla permite establecer un posible escenario de falla. Lo anterior hace relación con la capacidad de inferir acerca de la clase de elemento que está haciendo contacto con la línea en el lugar de falla. Esta clase de información puede ser relevante a la hora de determinar el tipo de procedimiento a seguir para solucionar el problema. El proceso de conformar dicha etapa, consiste en establecer grupos en los cuales se representan rangos de valores de resistencias de falla.

### ***EC3-Clasificación de los datos según el lugar de ocurrencia de la falla***

En esta etapa se utilizan los resultados obtenidos en las etapas anteriores, destacándose la información acerca del valor de resistencia de falla. Esta información se utiliza para establecer

límites entre los grupos y relaciones asociadas con la distancia de la falla, lo cual se traduce en un nuevo agrupamiento (zonificación) dentro de los rangos de valores de resistencia de falla. Lo anterior permite diferenciar entre una falla ocurrida en un punto alejado del sistema (alto valor de resistencia de falla) y una ocurrida en un punto cercano respecto al equipo de medida (bajo valor de resistencia de falla). Este modelo permite discriminar eventos que aparentemente se ubican en la misma zona, cuantificando la probabilidad de pertenencia a la misma.

#### 4.6.2. Validación del modelo

Se presentan los resultados de la validación, a partir de la cuantificación del número de aciertos y desaciertos obtenidos en cada una de las etapas del modelo. Estos resultados se organizan por tipos de falla. Posterior a los resultados se concluye acerca del desempeño del modelo, evaluando la capacidad de discriminación entre los datos y la obtención de grupos en cada una de las etapas. El proceso de validación y la forma de empleo del modelo se representa con el diagrama de flujo mostrado en la figura 4.10.

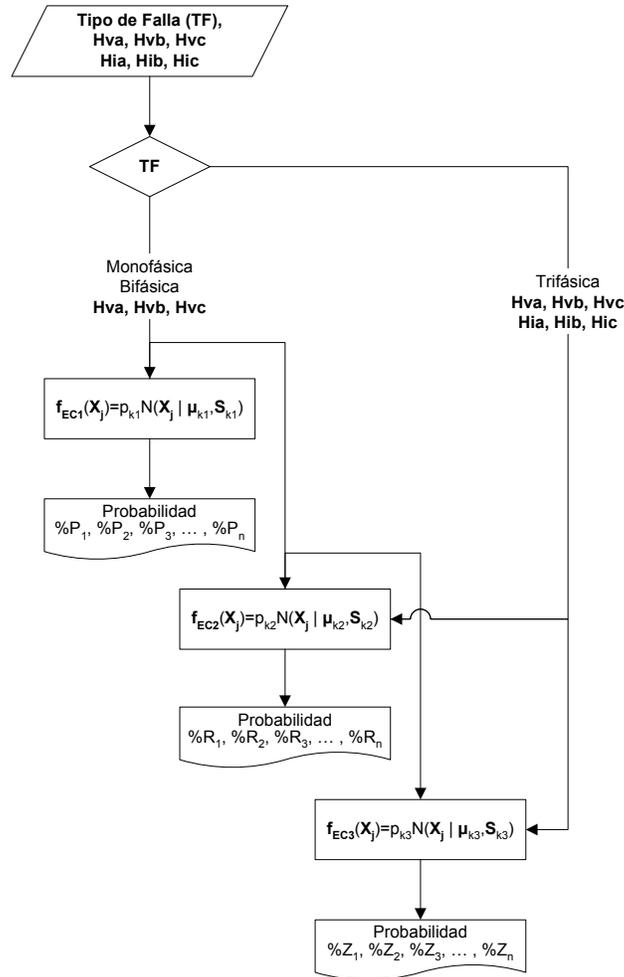


Figura 4.10: Diagrama de flujo de la validación y uso del modelo estadístico.



# Capítulo 5

## Pruebas y Resultados

### 5.1. Introducción

Este capítulo describe mediante un ejemplo de aplicación el entrenamiento y validación del modelo estadístico. Adicionalmente, se presentan los resultados obtenidos en la aplicación.

Cada falla registrada viene acompañada de información relevante a la hora de clasificar los datos dentro de una clase establecida en el modelo (zona dentro del sistema de distribución). La información utilizada en el estudio corresponde a magnitudes (no ángulos) de tensión y corriente, puesto que se quiere realizar una herramienta de bajo costo que no implique el empleo de equipos de última tecnología. Muchos de los equipos de medida que en la actualidad poseen las empresas, solo manejan las magnitudes de las señales. El sistema prototipo escogido se muestra en la figura 5.1.

### 5.2. Determinación del sistema y los datos a utilizar

Para la discriminación de las zonas se tuvo presente los grupos visualmente apreciables en el análisis preliminar de datos. El propósito de asociar los grupos a las zonas, es establecer la correspondencia de la ubicación de la falla en el sistema con la clasificación de los datos dentro de los grupos. Para representar la información del sistema se utilizaron tres descriptores que corresponden a la magnitud del hueco de tensión (caso monofásico y bifásico) y el aumento de la corriente (caso trifásico) en cada fase cuando la falla ocurre.

Para esta aplicación, los datos de entrenamiento y validación son simulaciones de fallas realizadas en cada barra del sistema con diferentes resistencias de falla. Los valores de resistencia de falla

utilizados en el estudio fueron tomados a partir del análisis realizado en [Das, 1998], [Dagenhart, 2000], [Uriarte, 2003], [Aucoin y Jones, 1996]. Allí se toma un modelo de resistencia de falla que tiene presente los efectos de la resistencia del arco eléctrico y la trayectoria a tierra. En la tabla 5.1 se presentan los valores de resistencia de falla utilizados para los procesos de entrenamiento (E) o validación (V).

Tabla 5.1: Valores de resistencia de falla para el proceso de entrenamiento y validación

<b>Resistencia de Falla</b> [ $\Omega$ ]	<b>Proceso</b>	<b>Resistencia de Falla</b> [ $\Omega$ ]	<b>Proceso</b>
0.05	E	25	E
5	E	30	V
10	V	35	E
15	E	40	V
20	V	50	E

En la tabla 5.2 se contabilizan el número de simulaciones realizadas para el entrenamiento (E) como para la validación (V), organizadas según el tipo de falla.

Tabla 5.2: Número de simulaciones realizadas para cada proceso

<b>Tipo de Falla</b>	<b>Simulaciones por Fase</b>			<b>Proceso</b>	
	<b>A</b>	<b>B</b>	<b>C</b>	<b>E</b>	<b>V</b>
Monofásica	132	187	176	315	180
	<b>AB</b>	<b>BC</b>	<b>CA</b>		
Bifásica	132	132	132	252	144
Bifásica-T	132	132	132	252	144
	<b>ABC</b>				
Trifásica	132			84	48
Trifásica-T	132			84	48
<b>Total</b>	<b>1551</b>			<b>1323</b>	<b>228</b>

## 5.3. Entrenamiento

Antes de iniciar la clasificación se realiza una verificación visual de los datos de entrenamiento con el fin de identificar y establecer previamente zonas dentro del sistema de distribución. El objetivo es estimar los centros iniciales de los grupos, ya que estos presentan influencia sobre la organización de los mismos por parte de los algoritmos a utilizar.

El número de grupos para la segunda y tercera etapa de clasificación se determinó partiendo del supuesto de que cada valor de resistencia de falla y cada barra del sistema es un posible grupo y además que cada una de las fallas de cada sección presenta diferente comportamiento respecto a las fallas de las secciones vecinas. Posteriormente y de acuerdo con la cantidad de comportamientos similares presentados en las diferentes barras del sistema, se conformaron grupos representativos a partir de la unión de varios grupos pequeños. Como consecuencia de este hecho se obtuvieron zonas de falla del sistema, asociadas a cada grupo representativo.

A continuación en la figura 5.1a se presenta cada uno de los grupos seleccionados posterior al entrenamiento para las fallas monofásicas. Se observan cuatro grandes grupos que contienen la totalidad de las barras del sistema. Mientras que para la fallas bifásicas y trifásicas se obtuvieron tres grupos representativos tal y como se muestra en la figura 5.1b.

En la tabla 5.3 se observa que nodos pertenecen a cada grupo conformado según el tipo de falla. Adicionalmente se destaca que las fallas bifásicas y trifásicas comparten las mismas zonas de falla, por consiguiente el sistema de distribución en estudio tendrá únicamente dos esquemas para las posibles zonas de falla, uno para zonas monofásicas y otro para las bifásicas y trifásicas, como se muestra en la figura 5.1.

Tabla 5.3: Barras que conforman cada una de las zonas de falla

Zona de Falla	Tipo de Falla	
	Monofásica	Bifásica y Trifásica
I	1,2,3,4	1,2,3,4
II	5,6,7,12	5,6,7,12
III	8,9,10,11	8,9,10,11
IV	13,14,15,16,17,18,19,20,21	No Presenta

Como se comentó en el capítulo 4, el modelo propuesto está estructurado por diferentes etapas

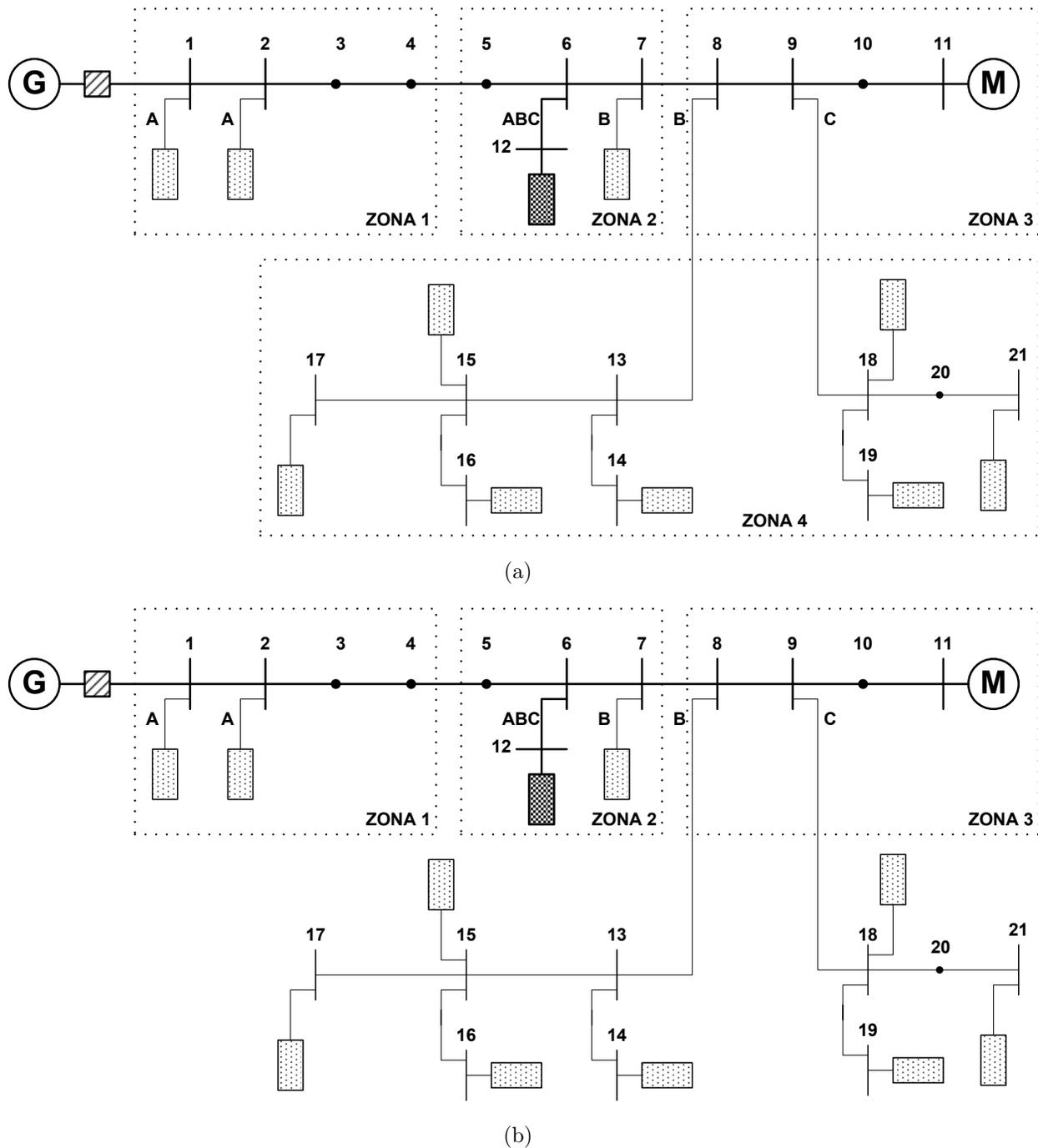


Figura 5.1: Zonas de falla del sistema. (a) Esquema para zonas monofásicas. (b) Esquema para zonas bifásicas y trifásicas.

de clasificación, la primera (EC1), determina la fase fallada, la segunda (EC2), determina el valor de resistencia de falla y la tercera (EC3), determina la zona del evento. La etapa EC1 aplica únicamente a las fallas monofásicas y bifásicas mientras que la EC2 y EC3 aplican a todas, debido a que el alimentador principal es trifásico mientras que los ramales externos son monofásicos. En

las figuras 5.2, 5.3 y 5.4 se muestra la distribución de los datos de entrenamiento correspondientes a tres tipos de falla, además se observan tres grupos claramente definidos, los cuales, corresponden a cada una de las fases en el caso monofásico y a las posibles combinaciones entre fases para el caso bifásico. Las gráficas fueron generadas con rutinas implementadas en MatLab.

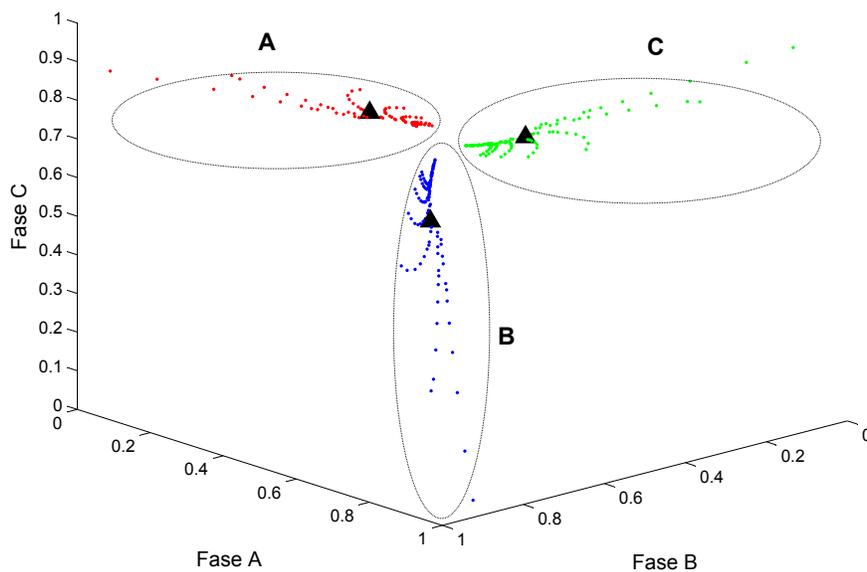


Figura 5.2: Distribución de los datos de entrenamiento falla monofásica.

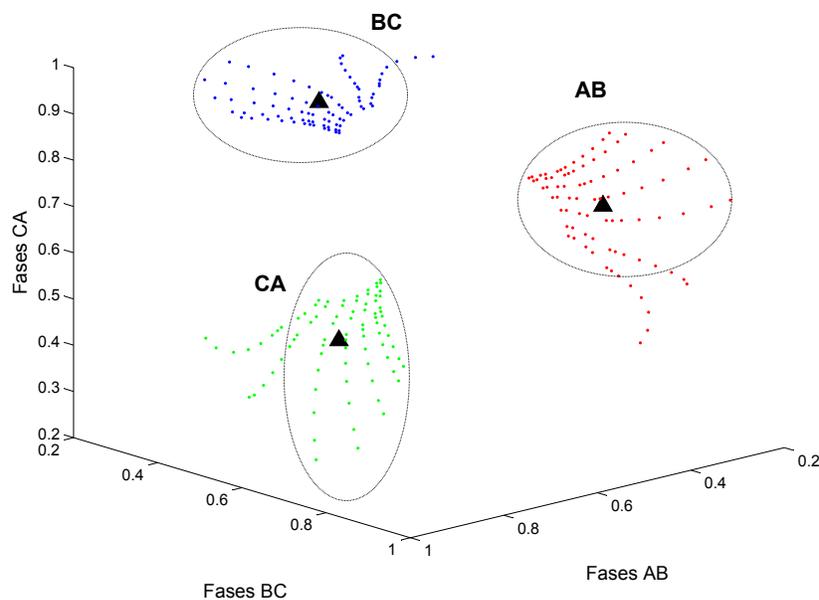


Figura 5.3: Distribución de los datos de entrenamiento falla bifásica.

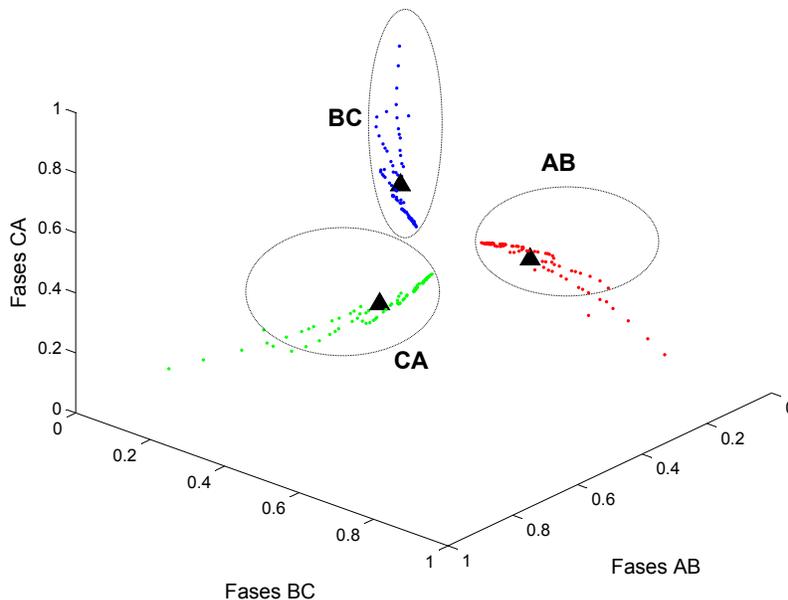


Figura 5.4: Distribución de los datos de entrenamiento falla bifásica a tierra.

A partir de la definición a priori de los grupos y con la aplicación del algoritmo k-medias se estiman los centros de los grupos con los cuales se inicia el análisis en la etapa EC1. Los centros se someten al criterio del analista para su escogencia con el fin de obtener los estimados más convenientes. Posteriormente, se define la forma y la proporción final en que los grupos se encuentran dentro de la distribución, utilizando el algoritmo EM y los estimados iniciales. En esta etapa, se utiliza un modelo heterocedástico, el cual permite determinar libremente la forma de las matrices de covarianzas para cada distribución y por ende la forma de cada grupo.

En la etapa EC2, la información acerca del valor de resistencia de falla permite crear un posible escenario donde se conforman grupos que representan rangos de valores de resistencias de falla. En este caso se establecieron cinco grupos representativos, tal y como se muestra en la figura 5.5. Lo anterior hace relación con la capacidad de predecir la causa de la falla. Esta clase de información puede ser relevante a la hora de determinar el tipo de procedimiento a seguir para solucionar el problema. En la figura 5.5, se muestra la forma y el tamaño final de los cinco grupos obtenidos (elipses). Los centros y los coeficientes de mezclado son representados por las estrellas. El valor de resistencia de falla aumenta, conforme se avanza hacia la derecha. La forma elíptica de cada grupo, está relacionada con los valores de la matriz de covarianzas.

En la etapa EC3 los datos se agrupan nuevamente. Cada grupo representa una zona probable de ocurrencia de la falla. Cada zona definida, está asociada a un número determinado de barras

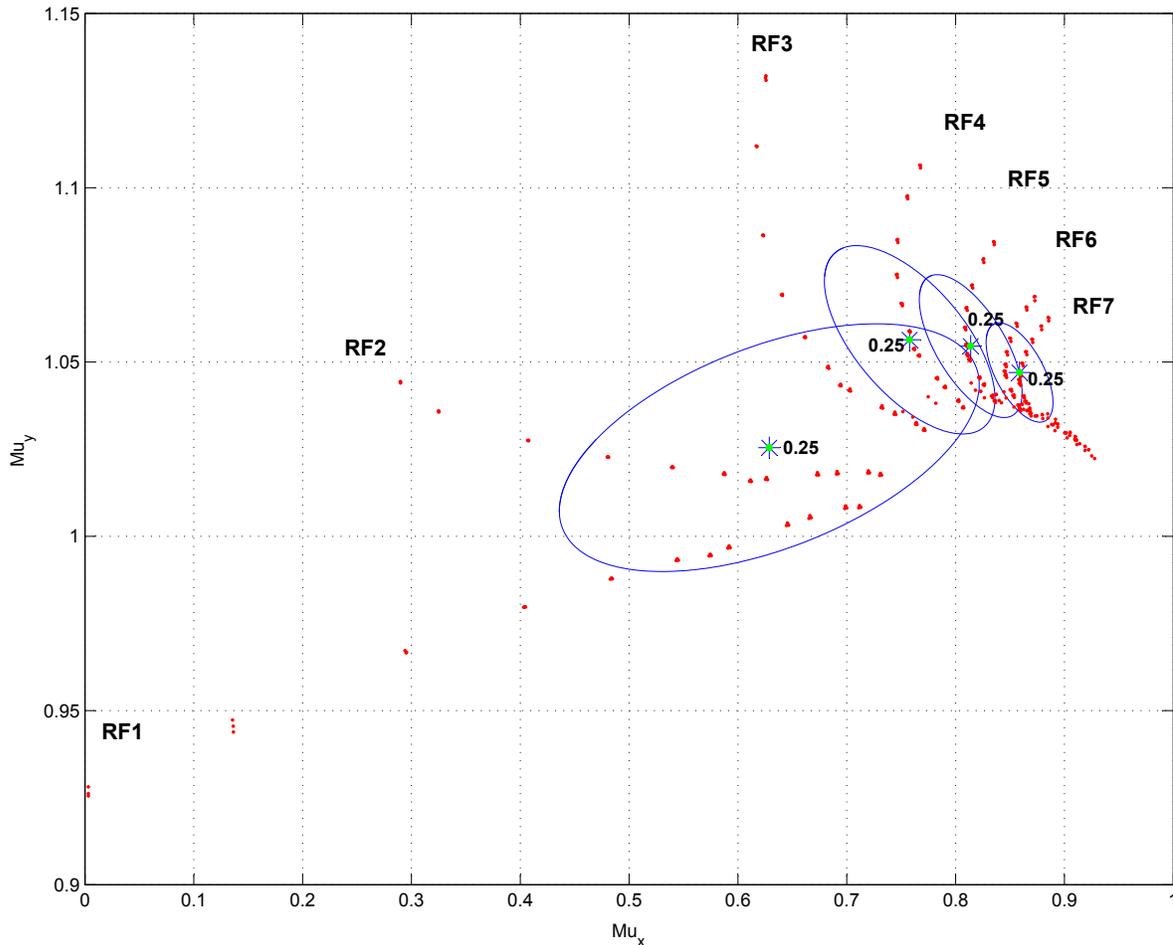


Figura 5.5: Distribución de los grupos según rangos de resistencia de falla para una falla monofásica (EC2).

dentro del sistema de distribución, tal y como se observa en la figura 5.6. Estas zonas se establecen con ayuda de la información acerca del número de grupos obtenido en la etapa EC2. Por lo tanto, en la etapa EC3 se tienen  $r$  grupos correspondientes a los grupos de resistencia de falla determinados en EC2 y estos a su vez contienen  $z$  grupos correspondientes a cada una de las zonas del sistema. En el caso del sistema prototipo, existe un modelo de mezclas compuesto por cuatro grupos (EC2) y cada grupo está compuesto por cuatro subgrupos que representan las zonas del sistema. En la figura 5.6 se ilustra la construcción de los grupos asociados a las zonas. Es de notar que los parámetros de cada grupo se muestran en la gráfica.

El algoritmo EM tiene la libertad de operar los elementos de la matriz de covarianza en cada

iteración hasta conseguir la mejor estimación. En este caso, si las matrices de covarianza permanecieran sin cambio en relación con su estimación inicial (matrices identidad) se obtendrían grupos de formas circulares. Finalmente, se aplica el modelo de mezcla, con el fin de comparar las zonas supuestas inicialmente (figuras 5.1a y 5.1b) con las determinadas por el algoritmo. En este caso de aplicación, la aproximación final de las zonas corresponde al resultado obtenido luego de la aplicación de los algoritmos. Se destaca, la influencia de una adecuada aproximación inicial de las zonas por parte del analista de los datos, permitiendo mejorar el ajuste del modelo de los datos durante su construcción.

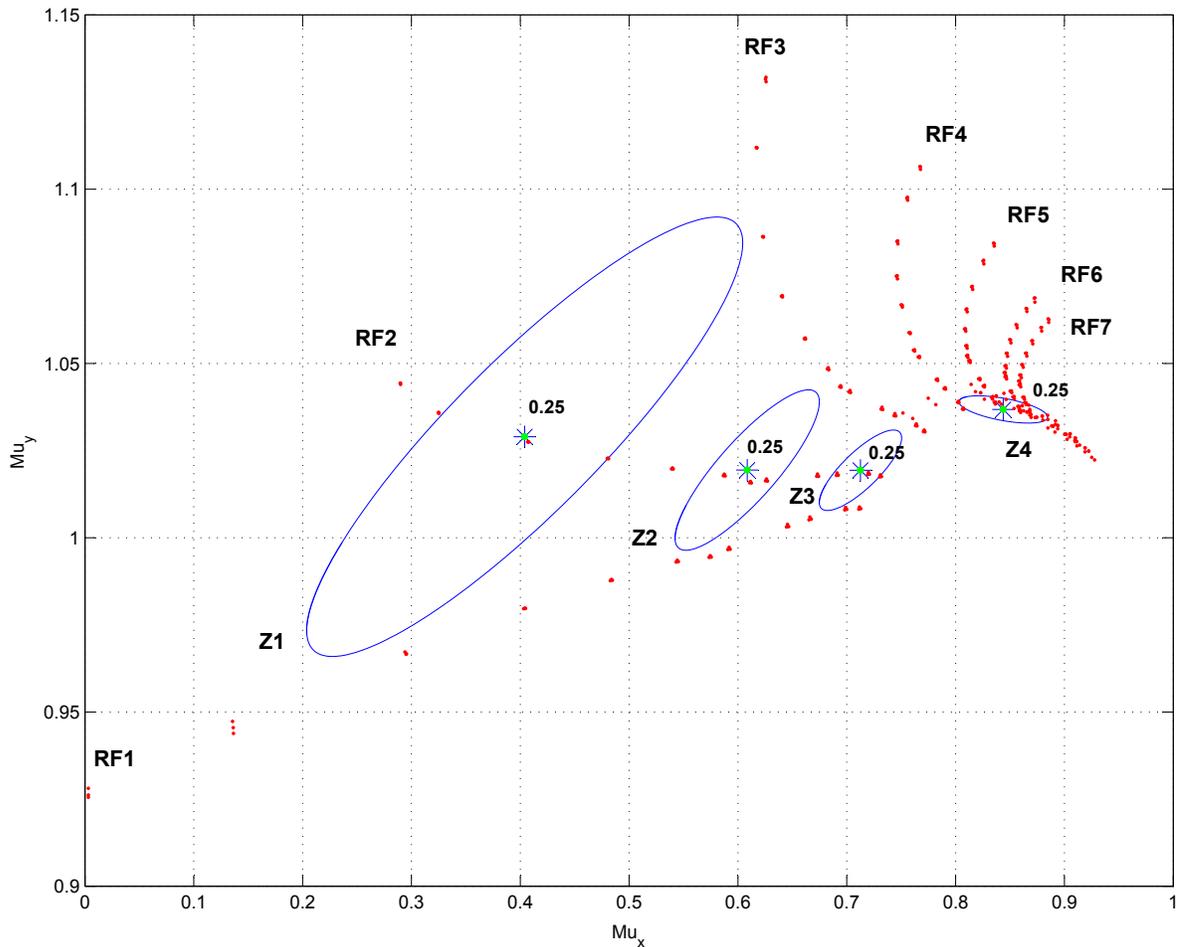


Figura 5.6: Distribución de los grupos (zonas) asociados al grupo 1 (rango 1) de la etapa EC3.

## 5.4. Validación

Establecido el modelo de los datos, se realiza la validación con datos desconocidos para el mismo. Las observaciones poseen datos con valores de resistencias de falla asociadas de 10, 20, 30 y 40  $\Omega$ . Los resultados de la validación se presentan como la relación entre la cantidad de aciertos y el número de datos de evaluados. Este valor también se presenta en forma de porcentaje de los datos evaluados que fueron clasificados acertadamente. Este valor se ubica seguido a la fracción y separado por dos puntos. Con el fin de validar la primera etapa de clasificación (EC1), se utiliza un conjunto de 468 observaciones.

En la tabla 5.4 se muestra el número de datos pertenecientes a fallas en las tres fases para el caso monofásico y los datos relacionados con las fallas entre fases. Es de resaltar que el modelo clasificó correctamente la totalidad de las observaciones pertenecientes a las fallas monofásicas y bifásicas, mientras que las bifásicas a tierra presentan un menor porcentaje de aciertos en la clasificación.

Tabla 5.4: Resultados de validación (Etapa EC1)

Tipo de falla	Fase / Fases			Subtotales
	A	B	C	
Monofásica	48/48 : 100,0	68/68 : 100,0	64/64 : 100,0	<b>180/180 : 100,0</b>
	AB	BC	CA	
Bifásica	48/48 : 100,0	48/48 : 100,0	48/48 : 100,0	<b>144/144 : 100,0</b>
Bifásica-T	33/48 : 68,75	34/48 : 70,83	34/48 : 70,83	<b>101/144 : 70,13</b>
			<b>Total</b>	<b>425/468 : 90,81</b>

De la misma forma se presentan los resultados obtenidos en la validación de los datos (564 observaciones) mediante el clasificador EC2. Es de notar que los porcentajes en la validación de esta etapa son relativamente altos, superiores al 80 % perteneciente a la falla monofásica, mientras que el máximo porcentaje lo presenta la falla trifásica con un 93,75 %. En tabla 5.5 se presenta el resumen de los resultados.

Los resultados presentados en la tabla 5.6, reflejan un buen desempeño del clasificador EC3 de acuerdo con el porcentaje de aciertos en la clasificación de los datos en las zonas definidas. Cabe anotar que se presentó el máximo número de aciertos en las fallas bifásicas y trifásicas, sin embargo los valores correspondientes a la falla bifásica a tierra presentan el número más bajo de aciertos, haciendo de este un modelo con bajo porcentaje de eficiencia para este tipo de fallas.

Tabla 5.5: Resultados de validación (Etapa EC2)

Tipo de falla	Rangos de RF				Subtotales
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	
Monofásica	27/45 : 60,00	36/45 : 80,00	36/45 : 80,00	45/45 : 100,0	<b>144/180 : 80,00</b>
Bifásica	27/36 : 75,00	30/36 : 83,33	36/36 : 100,0	36/36 : 100,0	<b>129/144 : 89,58</b>
Bifásica-T	24/36 : 66,66	30/36 : 83,33	36/36 : 100,0	36/36 : 100,0	<b>126/144 : 87,50</b>
Trifásica	20/24 : 83,33	24/24 : 100,0	24/24 : 100,0	24/24 : 100,0	<b>90/96 : 93,75</b>
				<b>Total</b>	<b>489/564 : 86,70</b>

Los resultados presentados en la tabla 5.6, reflejan un buen desempeño del clasificador EC3 de acuerdo con el porcentaje de aciertos en la clasificación de los datos en las zonas definidas. Cabe anotar que se presentó el máximo número de aciertos en las fallas bifásicas y trifásicas, sin embargo los valores correspondientes a la falla bifásica a tierra presentan el número más bajo de aciertos, haciendo de este un modelo con bajo porcentaje de eficiencia para este tipo de fallas.

Tabla 5.6: Resultados de validación (Etapa EC3)

Tipo de falla	Zonas				Subtotales
	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>	
Monofásica	48/48 : 100,0	48/48 : 100,0	48/48 : 100,0	22/36 : 61,11	<b>166/180 : 92,22</b>
Bifásica	48/48 : 100,0	48/48 : 100,0	48/48 : 100,0	No Presenta	<b>144/144 : 100,0</b>
Bifásica-T	22/48 : 45,83	32/48 : 66,66	6/48 : 12,5	No Presenta	<b>60/144 : 41,66</b>
Trifásica	32/32 : 100,0	32/32 : 100,0	32/32 : 100,0	No Presenta	<b>96/96 : 100,0</b>
				<b>Total</b>	<b>466/564 : 82,62</b>

Como un ejemplo, en la tabla 5.7 se presentan los resultados de la clasificación de cuatro observaciones, según las probabilidades calculadas en cada grupo. Cabe resaltar que este es un ejemplo para ilustrar el procedimiento debido a la gran cantidad de información procesada. Los datos mostrados corresponden a cuatro fallas monofásicas ubicadas en las cuatro zonas del sistema de la figura 5.1 y su correspondiente zona estimada por el modelo. A cada zona se le asigna una probabilidad de ocurrencia de la falla calculada en la etapa EC3. En consecuencia, las zonas estimadas están dadas por el mayor valor de probabilidad calculado en cada caso. Estos valores se encuentran resaltados en negrilla.

Tabla 5.7: Ejemplo de clasificación (Etapa EC3)

Zona Real	Probabilidad [%]				Zona Estimada
	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>	
1	<b>91,30</b>	8,69	0,01	0,00	1
2	0,77	<b>97,60</b>	1,59	0,04	2
3	0,17	9,17	<b>90,64</b>	0,02	3
4	0,10	0,47	4,38	<b>95,56</b>	4

## 5.5. Aplicación del algoritmo Ratan Das

A continuación se presentan los resultados obtenidos a partir de la aplicación del algoritmo Ratan Das para la localización de fallas en el sistema de distribución de la figura 5.1. Con el fin de facilitar la identificación y organización de los resultados, se utilizan identificadores numéricos en el momento de realizar las representaciones graficas. La corriente umbral utilizada fue de 70A, la cual fue obtenida a través de un análisis de cortocircuito y sugerida en [Das, 1998]. En la figura 5.7 se observa que el algoritmo presentó desaciertos en los tipos de falla trifásicas y monofásicas, mientras que en las fallas bifásicas y bifásicas a tierra mostró los mejores resultados.

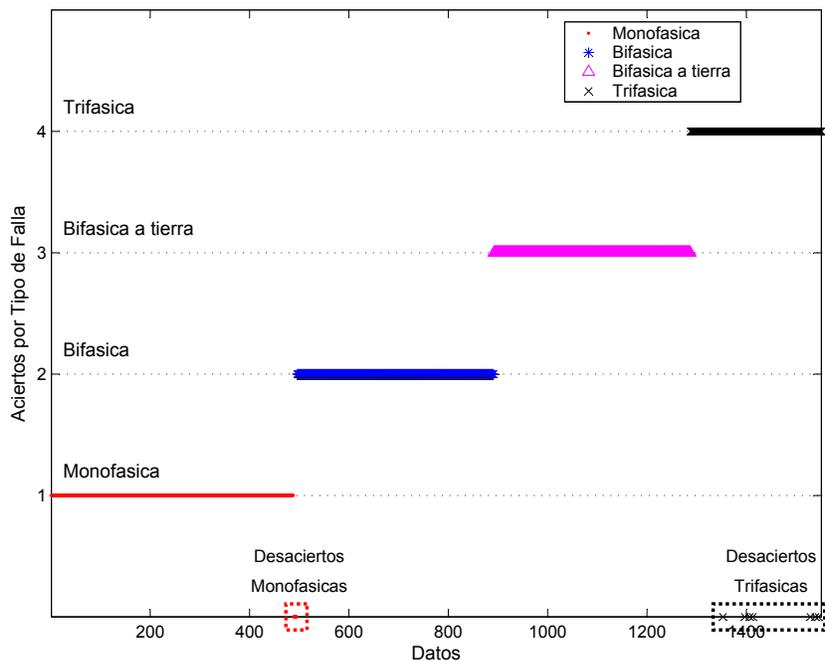


Figura 5.7: Resultados por Tipo de falla

En la tabla 5.8 se presentan los resultados de las estimaciones realizadas por el algoritmo. Estas son clasificadas según el tipo de falla. Se observa que el menor porcentaje de aciertos (97,34 %) se presenta en el tipo de falla trifásica, mientras que en total se obtuvo un 99,09 % de aciertos, lo cual indica que el algoritmo presenta buenos resultados al momento de identificar los tipos de falla.

Tabla 5.8: Resultados por Tipo de Falla

Id	Tipo de falla	Aciertos
1	Monofásica	488/495 : 98,58
2	Bifásica	396/396 : 100,0
3	Bifásica-T	396/396 : 100,0
4	Trifásica	257/264 : 97,34
<b>Total</b>		<b>1537/1551 : 99,09</b>

En la figura 5.8 se muestran los resultados de las estimaciones acerca del tipo de falla y la fase afectada. Se destacan los desaciertos obtenidos en las fallas monofásicas y trifásicas. Adicionalmente, se observa que las fallas bifásicas y bifásicas a tierra presentan aciertos en la totalidad de las estimaciones realizadas, dando cuenta de la efectividad del algoritmo para este tipo de fallas.

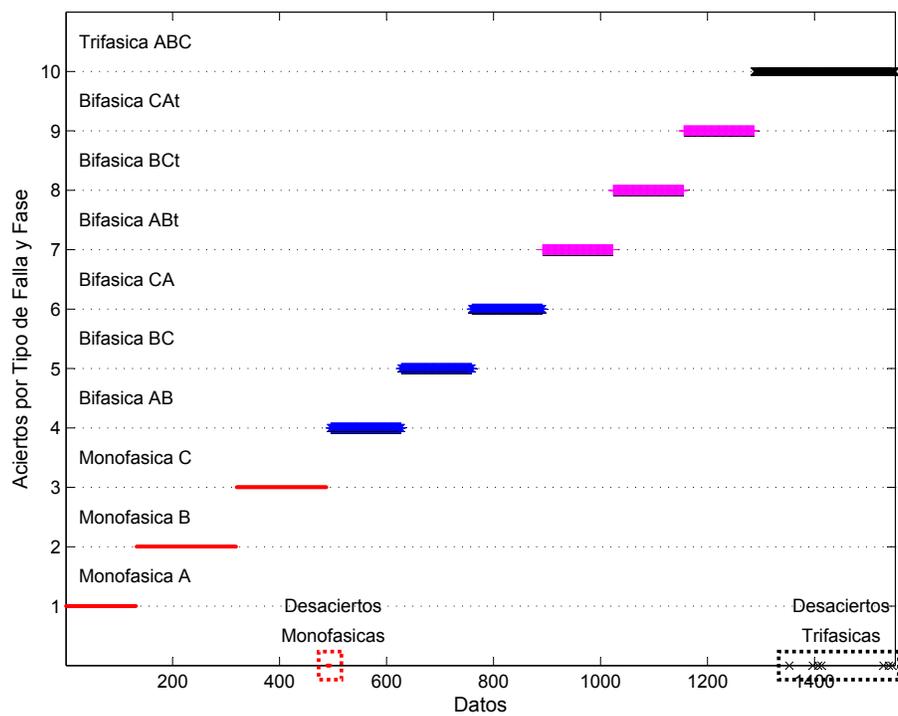


Figura 5.8: Resultados por Tipo de falla y fase Fallada

Los resultados de las estimaciones realizadas por el algoritmo mostrados en la tabla 5.9 se encuentran discriminados por el tipo de falla y las fases fallas. Se observa que los desaciertos se presentaron en las fallas monofásicas y trifásicas. En el primer caso, la fase *C* fue la única que presentó desaciertos en las fallas monofásicas. En el caso de las fallas trifásicas se observa que el porcentaje de desaciertos corresponden al 2,66 %.

Tabla 5.9: Resultados por Tipo de Falla y Fase

<b>Id</b>	<b>Tipo de falla</b>	<b>Fase/Fases</b>	<b>Aciertos</b>
1		A	132/132 : 100,0
2	Monofásica	B	187/187 : 100,0
3		C	169/176 : 96,02
4		AB	132/132 : 100,0
5	Bifásica	BC	132/132 : 100,0
6		CA	132/132 : 100,0
7		AB	132/132 : 100,0
8	Bifásica-T	BC	132/132 : 100,0
9		CA	132/132 : 100,0
10		ABC	257/264 : 97,34
<b>Total</b>	<b>Trifásica</b>		<b>1537/1551 : 99,09</b>

A continuación se analizan los resultados presentados por el algoritmo organizados según la cantidad de aciertos en los 21 nodos del sistema de distribución. En la tabla 5.10, se observa que el algoritmo acertó en cuanto a la ubicación de la falla en un 89,36 %. El mayor porcentaje de desaciertos se presentó en los nodos ubicados en el alimentador trifásico, este porcentaje corresponde a un 8,57 % del total de los datos, mientras que los ramales monofásicos presentan un 2,07 % de desaciertos respecto al total. Cabe resaltar que si se cuantifica la cantidad de desaciertos en los ramales monofásicos de forma separada, estos representan un 67,67 % del total de datos de los ramales. En el caso del alimentador trifásico este mismo razonamiento representa un 9,16 % de los datos.

En la figura 5.9 se presenta la cantidad de aciertos obtenidos tanto por tipo de falla como por fase en cada nodo del sistema. Como se muestra en la figura 5.9, en el nodo 1, el algoritmo acertó en un 100 % en cuanto al tipo de falla, es decir, el algoritmo identificó correctamente la totalidad las fallas monofásicas, bifásicas, bifásicas a tierra y trifásicas, producidas en el nodo 1, sin embargo, no fue tan eficaz en las estimaciones de la ubicación de la falla mostrando un porcentaje del 90,08 % en cuanto a la estimación del nodo. Se observa de forma comparativa que el algoritmo presenta

mayores porcentajes de acierto en la identificación de la falla que en la ubicación de la misma en varios nodos del sistema. Este resultado se repite sin de forma similar tanto en el alimentador trifásico como en el monofásico.

Tabla 5.10: Resultados por Nodos

<b>Nodo</b>	<b>Aciertos</b>	<b>Nodo</b>	<b>Aciertos</b>
1	109/121 : 90,08	12	55/121 : 45,45
2	121/121 : 100,0	13	11/11 : 100,0
3	121/121 : 100,0	14	7/11 : 63,63
4	121/121 : 100,0	15	11/11 : 100,0
5	121/121 : 100,0	16	3/11 : 27,27
6	120/121 : 99,17	17	3/11 : 27,27
7	121/121 : 100,0	18	11/11 : 100,0
8	121/121 : 100,0	19	6/11 : 54,54
9	121/121 : 100,0	20	11/11 : 100,0
10	118/121 : 97,52	21	4/11 : 36,36
11	70/121 : 57,85	<b>Total</b>	<b>1386/1551 : 89,36</b>

Se observa que los nodos 11 y 12 se presentaron la menor cantidad de aciertos en la localización respecto a los demás nodos pertenecientes al alimentador trifásico, dichos porcentajes son del 57,85 % y 45,45 % respectivamente. De los nodos correspondientes a los ramales monofásicos los que tuvieron el menor porcentaje de aciertos en cuanto a la ubicación de la falla fueron los nodos 16, 17 y 21 esto como consecuencia de la mala estimación del algoritmo debido a la incidencia de la distancia y los ramales asociados. El porcentaje correspondiente a los nodos 16 y 17 es del 27,27 %, mientras que para el nodo 21 es del 36,36 %.

De forma similar a la identificación del tipo de falla se observa la correcta identificación de las fases falladas en la mayoría de nodos asociados al alimentador trifásico. Se resaltan los porcentajes de acierto en los nodos 10, 11 y 12, los cuales equivalen al 98,34 % respectivamente. En el nodo 21 se observa un porcentaje del 36,36 % en los aciertos obtenidos. Se observa que los nodos con la menor cantidad de aciertos se encuentran ubicados al final del alimentador trifásico (nodo 10 y 11), así como el nodo 21 se encuentra ubicado al final del ramal monofásico. Adicionalmente, se resalta la ubicación del nodo 12 en una ramificación del alimentador principal. Los nodos finales en cada alimentador o ramal presentan problemas en la estimación, debido a la convergencia del algoritmo en ese punto [Villamizar y Quiñones, 2005].

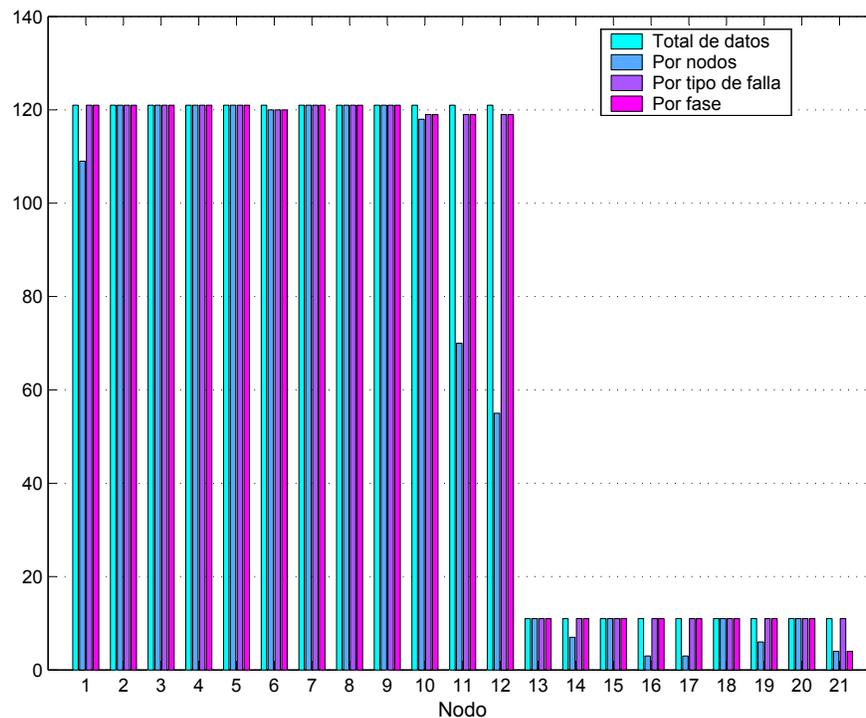


Figura 5.9: Cantidad de aciertos por nodo y por tipo de falla

## 5.6. Aplicación del híbrido Ratan Das - Modelo de Mezclas

En esta sección se realizan pruebas a partir de la combinación entre las zonas obtenidas con el modelo de mezclas y las múltiples estimaciones obtenidas con el método algorítmico de Ratan Das. Los resultados arrojados se analizan con el fin de evaluar esta alternativa para mejorar la localización de fallas en el sistema. Se analizaron los resultados de la prueba, determinando cuatro tipos de respuesta asociadas a los posibles casos con los cuales se describe el proceso de localización. Los casos analizados se describen de forma grafica en la figura 5.10.

- **Caso 1** (*Intersección Completa*) La totalidad de las multiples estimaciones arrojadas por el algoritmo Ratan Das, se encuentran dentro de la zona de falla arrojada por el modelo de mezclas.
- **Caso 2** (*Intersección Media*) Algunas de las estimaciones se encuentran dentro de la zona y otras se encuentran fuera de la misma.
- **Caso 3** (*Intersección vacía*) Ninguna de las estimaciones se encuentra dentro de la zona.
- **Caso 4** (*No Convergencia*) Por razones de convergencia, el algoritmo Ratan Das no puede

realizar estimaciones. Solo se tiene como salida del modelo híbrido la zona de falla.

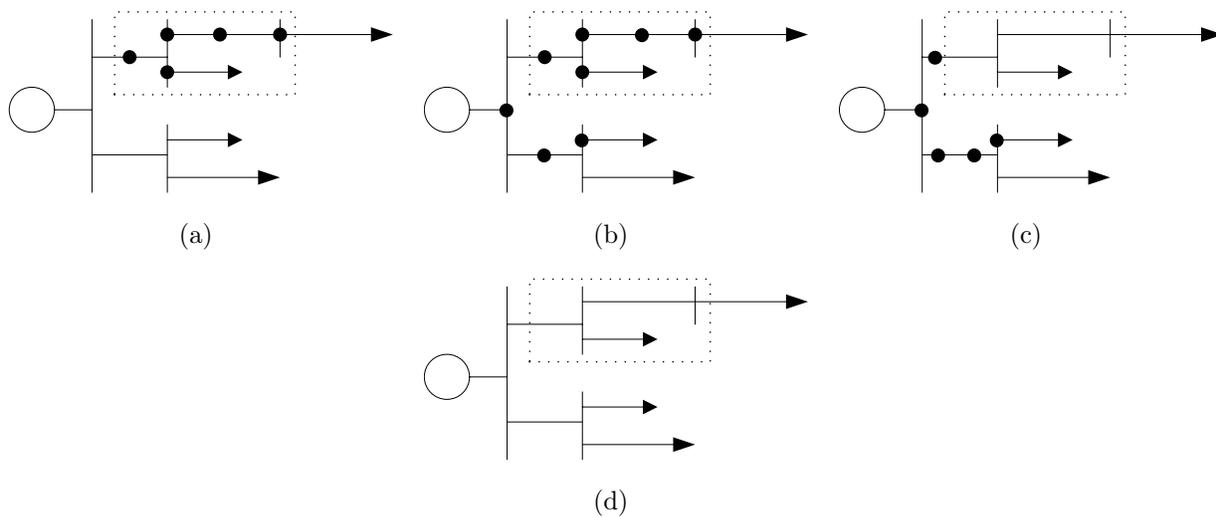


Figura 5.10: Resultados del modelo híbrido. (a) Caso 1. (b) Caso 2. (c) Caso 3. (d) Caso 4.

En la figura 5.11 se muestran los resultados arrojados por el modelo híbrido según cada caso. Los resultados se encuentran organizados por tipos de falla.

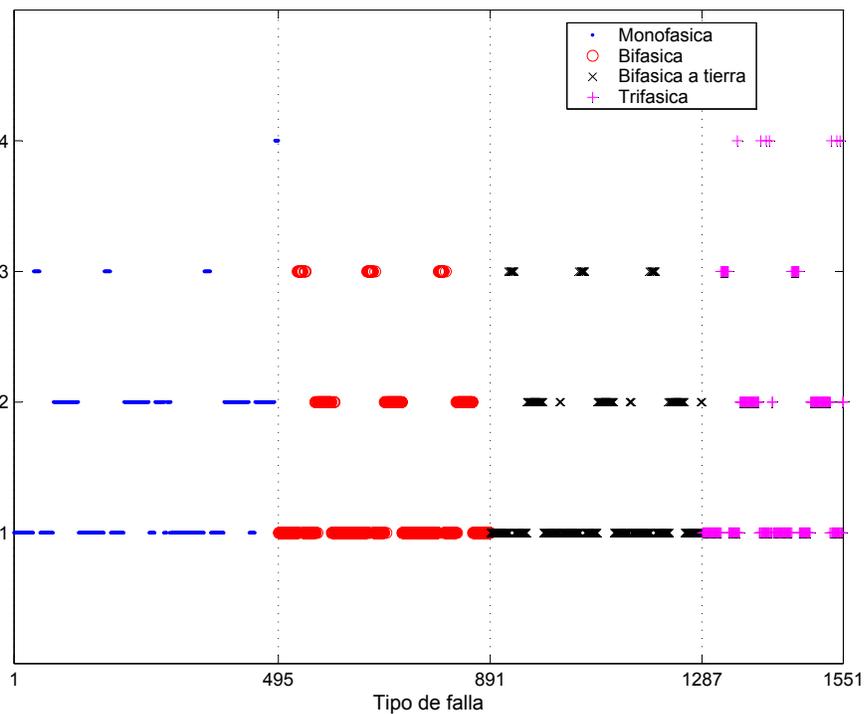


Figura 5.11: Resultados del modelo híbrido para los tipos de falla evaluados

De forma general se observa la presencia de los diferentes casos en los tipos falla monofásica y trifásica, mientras que en las fallas bifásicas y bifásicas a tierra no se presenta el caso 4. Se resalta que la mayor cantidad de los resultados se ubican en el caso 1 y la menor en el caso 4. La existencia del caso 4, muestra la incapacidad del algoritmo Ratan Das para estimar una solución debido a la no convergencia. En este caso el resultado es la zona estimada. El efecto de que la mayor cantidad de los resultados se encuentran ubicados en el caso 1, se puede interpretar como alta eficiencia del modelo híbrido ya que tanto las estimaciones como la zona se interceptan de forma completa, es decir, la falla se encuentra localizada dentro de la zona estimada y la estimación presenta alta probabilidad de acierto. No obstante, la ubicación de una cantidad representativa de los resultados en el caso 2, reduce la cantidad de estimaciones realizadas por el algoritmo Ratan Das, permitiendo dar una alternativa de solución al problema de la múltiple estimación, presente en los algoritmos de localización de fallas para sistemas de distribución. Finalmente, en el caso 3 se observa que no existe intersección entre los resultados arrojados por el modelo estadístico y el algoritmo, lo cual deja una incertidumbre acerca de cual de las dos opciones se debe escoger.

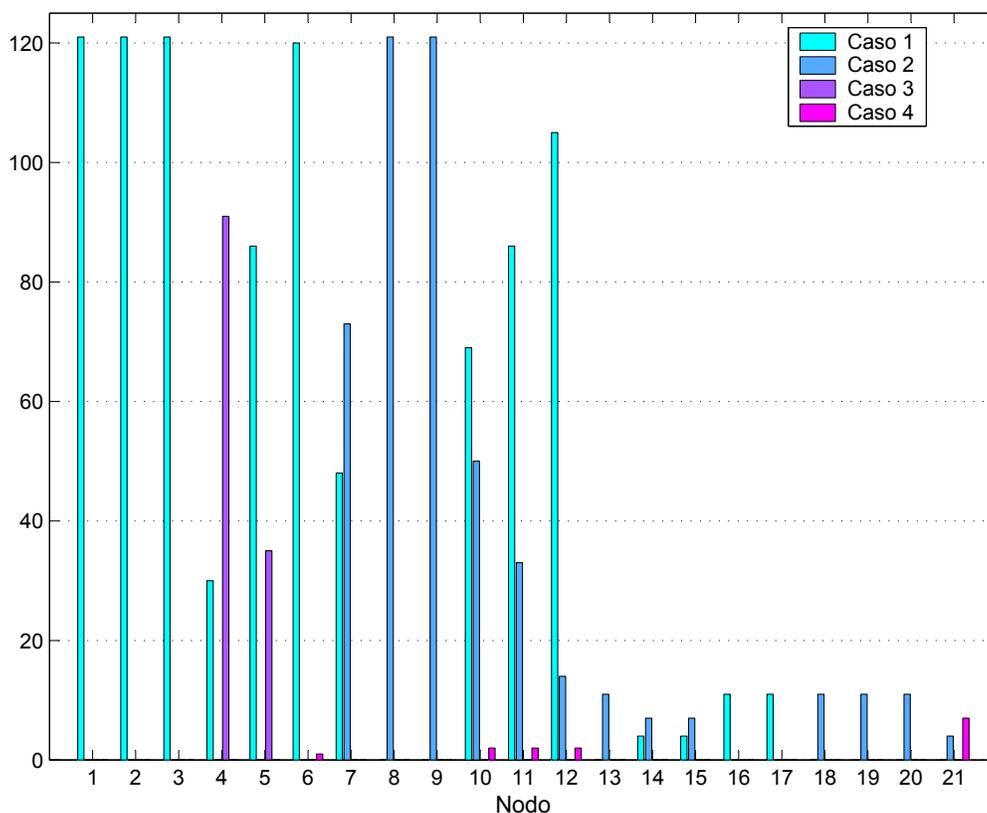


Figura 5.12: Resultados del modelo híbrido para cada nodo del sistema

En la figura 5.12 se presenta la distribución de los resultados del modelo híbrido en cada uno de los nodos del sistema. Los resultados del modelo se analizan de forma similar que en el tópico anterior, destacando la presencia de los diferentes casos en los nodos del sistema.

En relación a los nodos pertenecientes al alimentador trifásico se puede resaltar que los resultados ubicados en el caso 4 se presentaron en los nodos 6 y 12, donde el alimentador trifásico posee una derivación trifásica y en los nodos 11 y 12, que son los nodos finales del alimentador. En estos nodos, el algoritmo presentó confusión y no pudo realizar las estimaciones necesarias. En el caso del ramal monofásico, el nodo 21 fue el único que no obtuvo estimación. Es de resaltar que el algoritmo no presentó resultados en los nodos finales de cada ramal y en donde se encuentran ubicadas las derivaciones. La interpretación de los resultados de los nodos 4 y 5 donde se presentó el caso 3, se puede orientar hacia la diferencia en las estimaciones realizadas. Los casos 1 y 2 presentan una distribución esperada en la mayoría de nodos.

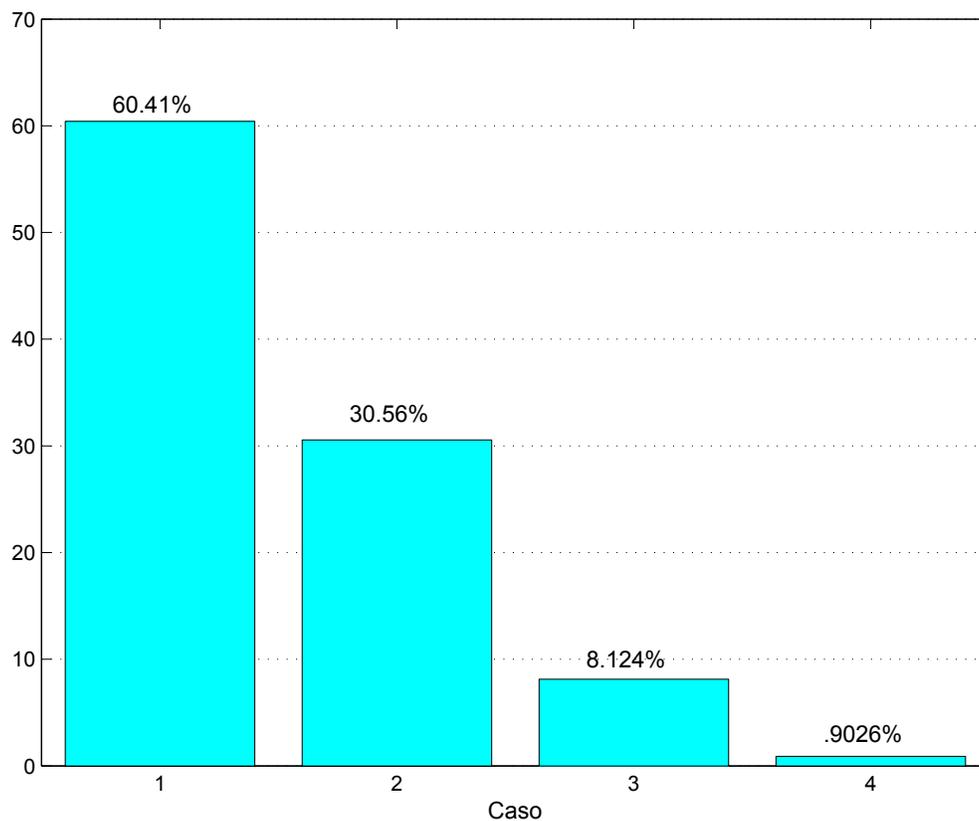


Figura 5.13: Porcentaje de los resultados distribuidos por caso

En la figura 5.13, se muestra los porcentajes de los cuatro casos respecto al número total de simulaciones realizadas. Tal y como se comentó anteriormente, existe una gran porcentaje de efectividad en el modelo, el cual es representado como la suma de los porcentajes de los casos 1 y 2. Este valor corresponde al 90,97 %, mientras que el porcentaje de error del modelo está asociado a los casos 3 y 4, para el caso más desfavorable, lo cual corresponde al 9,03 % de la totalidad de los datos.

Las proporciones entre los diferentes casos fue la esperada ya que se quiere que el modelo híbrido permita mejorar los resultados arrojados por ambos métodos. Los resultados se pueden corroborar resaltando que se pudo reducir el número de estimaciones arrojado por el algoritmo, lo cual se ve reflejado en el caso 2 con un porcentaje del 30,56 %. Seguidamente, se observa que el modelo híbrido mejora la eficiencia en los resultados debido a que entrega una zona y las posibles ubicaciones de la falla dentro y fuera de esta (caso 1 y 2). Adicionalmente, presenta un margen de incertidumbre del 8,12 % dado que se presenta el caso 3. Finalmente, se observa que el modelo siempre arrojará resultados aunque el algoritmo no estime solución alguna (caso 4).

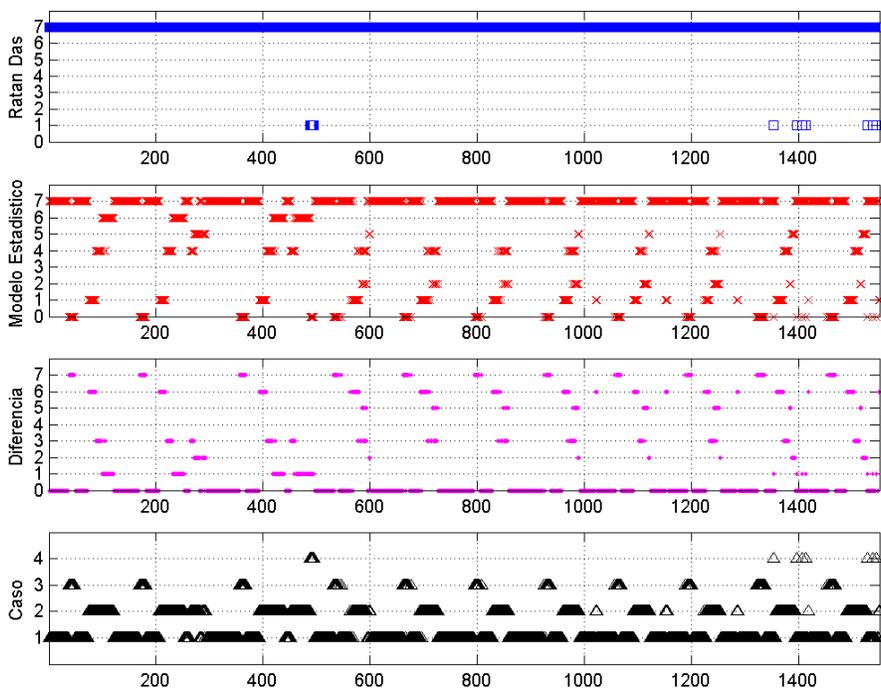


Figura 5.14: Comparativo entre las estimaciones realizadas por el algoritmo Ratan Das y las reducidas por el modelo estadístico

En la figura 5.14 se muestra el número de estimaciones arrojadas por el algoritmo Ratan Das (azul), así como el número de estimaciones reducidas por el modelo estadístico (rojo), la diferencias entre los resultados de ambos métodos (magenta) y el caso al cual pertenece la estimación (negro).

Este análisis se realiza con el objetivo de mostrar la reducción de la múltiple estimación que presenta el algoritmo al combinarlo con el modelo estadístico. Como se mencionó anteriormente los casos asociados a este objetivo son el 1 y 2. Es de notar que el valor de la diferencia entre las estimaciones debe ser menor que el valor arrojado por el algoritmo, lo cual refleja la motivación de este análisis. De otro lado, los casos 3 y 4 se encuentran representados por ceros en el diagrama correspondiente a las estimaciones reducidas por el modelo estadístico. Para estos dos casos los valores correspondientes en el diagrama de diferencias es siete ya que no presenta reducción alguna luego de la combinación de los dos métodos.

# Capítulo 6

## Estimación de los índices de calidad del servicio DES y FES

### 6.1. Introducción

Este capítulo presenta una metodología que permite estimar los índices DES y FES a partir de la cuantificación de los efectos y actividades asociadas a la presencia de una interrupción en el sistema.

La función de las empresas del sector eléctrico es generar, transportar, distribuir y comercializar energía eléctrica en forma segura, a un costo razonable y con un nivel de calidad adecuado a los requerimientos de los usuarios. Los aspectos técnicos que componen la calidad del servicio eléctrico son la calidad de la onda y la continuidad del suministro. En Colombia, este último se cuantifica mediante los índices de calidad DES y FES. El suministro de energía eléctrica no sólo se considera una necesidad básica, sino un producto que cobra importancia debido al impacto económico sobre los procesos, la producción y el bienestar de los usuarios.

Las interrupciones del suministro pueden surgir como consecuencia de fallas sobre las redes y en los equipos, por lo tanto, debido a los riesgos existentes, las fallas e interrupciones aparecen como inevitables, generando pérdidas económicas en los procesos y la producción. Adicionalmente se ve afectado el bienestar de los usuarios, quienes requieren de este servicio para realizar muchas de sus actividades cotidianas. Por esta razón, es necesario contemplar estrategias para la disminución del impacto de las interrupciones garantizando niveles de confiabilidad aceptables a bajo costo [Beltrán y Romero, 2006].

## 6.2. Índices de calidad del servicio

La resolución CREG070 de 1998 presenta los indicadores para la medición de la calidad del servicio y establece definiciones únicas para esos indicadores. Las expresiones son (6.1) y (6.2).

### Duración Equivalente de las Interrupciones del Servicio -DES-

$$DES_c = \sum_{i=1}^{NTI} t(i) \quad (6.1)$$

- **DESc** Sumatoria del tiempo en horas de las interrupciones del servicio en un circuito. Se calcula anualmente.
- **i** Interrupción  $i$ -ésima.
- **t(i)** Tiempo en horas de la interrupción  $i$ -ésima.
- **NTI** Número total de interrupciones ocurridas en el circuito. Se calcula anualmente.

### Frecuencia Equivalente de las Interrupciones del Servicio -FES-

$$FES_n = NTI \quad (6.2)$$

- **FESn** Frecuencia promedio por usuario, de las interrupciones del servicio en el nivel de tensión  $n$ , durante los últimos 12 meses.

## 6.3. Determinación de los tiempos de indisponibilidad

El procedimiento consiste en la división del tiempo total de la interrupción en intervalos de tiempo asociados a las actividades realizadas en los procedimientos de atención y recuperación de fallas [Carrillo et al., 1993], tal y como se muestra en la figura 6.1.

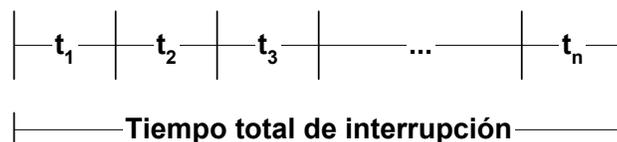


Figura 6.1: Intervalos del tiempo total de interrupción.

Para establecer la influencia de la información de localización en los procedimientos de atención y recuperación de fallas, se toma la duración total de una interrupción y se divide en intervalos de tiempo asociados a las acciones realizadas durante la reparación. Dado que no todos los operadores de red y empresas distribuidoras<sup>1</sup> tienen implementados los mismos procedimientos para la atención de fallas, se deben plantear intervalos de tiempo generales que permitan abarcar la mayor parte de las actividades de restauración del servicio. Se debe tener presente que esta división es flexible y que al momento de implementar la metodología, se pueden agregar o eliminar algunas actividades, según su conveniencia. De acuerdo con lo mencionado anteriormente, y habiendo revisado diferentes estudios sobre recuperación de fallas, se plantearon los siguiente intervalos de tiempo [Cárdenas y Olaya, 2002], [CREG, 2004].

1. ***Tiempo de detección (Td)*** Tiempo que tarda el OR en reconocer la existencia de una falla.
2. ***Tiempo de estimación del lugar de la falla (Te)*** Tiempo necesario para encontrar el área donde se estima ocurrió la falla.
3. ***Tiempo de planeación (Tp)*** Tiempo transcurrido entre la confirmación de existencia de falla y la emisión de la orden de trabajo.
4. ***Tiempo de traslado (Tt)*** Tiempo utilizado para llegar a la zona de falla.
5. ***Tiempo de localización (Tl)*** Tiempo necesario para realizar pruebas en el circuito o recorridos con el fin de determinar el lugar exacto de la falla.
6. ***Tiempo de seccionamiento (Tsc)*** Tiempo empleado en aislar el lugar de la falla.
7. ***Tiempo de maniobra para la transferencia (Tmt)*** Tiempo necesario para conectar las suplencias cuando se cuenta con estas y restablecer el servicio a los tramos en donde sea posible.
8. ***Tiempo de seguridad (Ts)*** Tiempo requerido para instalar las puestas a tierra y los equipos de seguridad.
9. ***Tiempo de logística (Tlg)*** Tiempo necesario para el despacho y transporte de materiales y equipos requeridos al lugar de la falla.
10. ***Tiempo de reparación y montaje (Trm)*** Tiempo para la reparación o reemplazo y montaje de los elementos.

---

<sup>1</sup>De aquí en adelante se utilizará la sigla **OR** para referirse a los operadores de red y empresas distribuidoras.

11. *Tiempo de desconexión de seguridad (Tds)* Tiempo para retirar las puestas a tierra instaladas.
12. *Tiempo de prueba (Tpr)* Tiempo para realizar las pruebas de funcionamiento y revisión final del trabajo.
13. *Tiempo de restablecimiento (Trs)* Tiempo empleado para restablecer el servicio en la parte aislada de la línea.
14. *Tiempo de desconexión de suplencias (Tv)* Tiempo necesario para recuperar la configuración normal de los tramos alimentados por suplencias, una vez ejecutadas las tareas de reparación en el tramo fallado.

## 6.4. Método para la estimación de los índices de calidad

Tras haber realizado la selección del circuito a analizar e identificar cada uno de sus tramos y demás elementos que lo componen, se procede a evaluar el funcionamiento del circuito ante la presencia de fallas en cada uno de los tramos, para esto se re realiza el siguiente procedimiento [Arriaga, 1994], [Beltrán y Romero, 2006].

### 6.4.1. Determinación del estado de los tramos

En función de la protección asociada, las alternativas de alimentación y la existencia de una falla en otro tramo del alimentador o en sí mismo, cada tramo del sistema tendrá un comportamiento, donde  $i$  representa el tramo a evaluar y  $j$  el tramo fallado, este comportamiento se puede definir de la siguiente forma:

- **Normal (N)** El estado del tramo del alimentador  $i$  se define como NORMAL, cuando su operación no se ve afectada por fallas en el tramo  $j$ .
- **Restablecible (R)** El estado del tramo de alimentador  $i$  es RESTABLECIBLE, cuando su servicio puede retornar a la normalidad sin ser desconectado de la alimentación principal y antes de reparar el tramo  $j$  fallado (aislando  $j$  mediante algún elemento de maniobra).
- **Transferible (T)** El tramo del alimentador  $i$  será TRANSFERIBLE, cuando exista una maniobra para re-energizarlo, antes de reparar el tramo fallado  $j$ .
- **Irrestablecible (I)** Son tramos IRRESTABLECIBLES aquellos que sufren la falla y todos los que no se pueden restablecer.

- **Irrestablecible con espera (IE)** El tramo  $j$ , en falla y los tramos que no pueden ser restablecidos, se definen como IRRESTABLECIBLES CON ESPERA, cuando previo a las reparaciones se debe realizar alguna maniobra de transferencia.

Para organizar la información se construye una matriz de estados, tal y como se muestra en la tabla 6.1, donde las filas representan el tramo fallado y las columnas representan la condición del tramo ante una falla en el sistema. Se llena cada casilla con las letras que representan cada estado. La matriz de estados es de orden  $(n \times n)$ , donde  $n$  es el número de tramos de la red. El procedimiento de construcción de la matriz de estados se puede sintetizar en tres pasos, el primero, definir el número de tramos de la red, el segundo, suponer una falla en uno de los tramos y finalmente el tercero, analizar el efecto de la falla en los demás tramos del sistema y determinar su estado.

Tabla 6.1: Matriz de estados

Tramo	1	2	...	n
1	$E_{11}$	$E_{12}$	...	$E_{1n}$
2	$E_{21}$	$E_{22}$	...	$E_{2n}$
⋮	⋮	⋮		⋮
n	$E_{n1}$	$E_{n2}$	...	$E_{nn}$

### 6.4.2. Cálculo del tiempo de interrupción de los tramos

Una vez definida la matriz se estados se procede a calcular el tiempo de interrupción total de cada tramo. Éste se obtiene sumando los tiempos asociados a cada estado, según se indica en la tabla 6.2.

Tabla 6.2: Tiempos de interrupción

Estado	Tiempo de interrupción
Normal	0
Restablecible	$T_d + T_e + T_p + T_t + T_l + T_{sc}$
Transferible	$T_d + T_e + T_p + T_t + T_l + T_{sc} + T_{mt} + T_v$
Irrestablecible	$T_d + T_e + T_p + T_t + T_l + T_{sc} + T_s + T_{lg} + T_{rm} + T_{ds} + T_{pr} + T_{rs}$
Irrestablecible con espera	$T_a + T_e + T_p + T_t + T_l + T_{sc} + T_{mt} + T_s + T_{lg} + T_{rm} + T_{ds} + T_{pr} + T_{rs}$

Los intervalos de tiempo asociados a los estados varían de acuerdo con la división de tiempos establecida por el OR y teniendo presente la definición de cada estado. Los tiempos de interrupción

se reemplazan en la matriz de estados para obtener la matriz de tiempos de interrupción que se muestra en la tabla 6.3.

Tabla 6.3: Matriz de tiempos de interrupción por tramos

Tramo	1	2	...	n
1	$r_{11}$	$r_{12}$	...	$r_{1n}$
2	$r_{21}$	$r_{22}$	...	$r_{2n}$
⋮	⋮	⋮		⋮
n	$r_{n1}$	$r_{n2}$	...	$r_{nn}$

### 6.4.3. Cálculo de la indisponibilidad por tramo

Para ello se debe caracterizar cada tramo de alimentador por su tasa de falla ( $\lambda$ ), la cual representa la cantidad de veces que un usuario se someten a interrupciones del suministro, por unidad de tiempo. Para un tramo, la tasa de falla indica el promedio de las veces que dicha sección se ve sometida a alguna condición que provoque una interrupción del servicio. Este parámetro se puede determinar a través del historial de fallas para el tramo individual ó mediante una estimación, considerando el sistema completo. La estimación se obtiene a partir de las expresiones (6.3) y (6.4).

$$b = \frac{m}{L \cdot T} \left[ \frac{Fallas}{km \cdot año} \right] \quad (6.3)$$

$$\lambda = b \cdot l \left[ \frac{Fallas}{año} \right] \quad (6.4)$$

Donde,  $m$  corresponde a la cantidad de fallas,  $L$  es la longitud total de los tramos de líneas expuestos a falla (kilómetros),  $T$  periodo de estudio (años),  $b$  número de fallas por kilómetro por año y  $l$  la longitud del tramo de línea de interés.

Para obtener la cantidad de interrupciones es necesario tener en cuenta que cada tramo presenta cierta cantidad de fallas, pero el número total de cortes es mayor debido a que se deben considerar los efectos de las fallas en los otros tramos. Dependiendo del estado definido para cada tramo se registra el número de interrupciones indicado en la tabla 6.4 en la matriz de interrupciones de la tabla 6.5. Para tres de los estados solo es necesario realizar una interrupción del servicio, a diferencia del número de interrupciones en el estado NORMAL y TRANSFERIBLE. En el primer caso, el tramo no se ve afectado de ninguna forma por la presencia de una falla, en consecuencia no

presenta ningún tipo de corte. Cuando un tramo es transferible, aparece dos veces la tasa de falla, ya que luego de realizar la reparación del tramo fallado y habiéndose presentado un corte inicial del servicio, se tiene que regresar a la configuración normal del sistema, por lo tanto es necesario interrumpir el servicio nuevamente para realizar la desconexión de la suplencia durante un tiempo  $T_v$ .

Tabla 6.4: Interrupciones aportadas según estado

Estado	Interrupciones $[\lambda]$
Normal	0
Restablecible	$\lambda$
Transferible	$2\lambda$
Irrestablecible	$\lambda$
Irrestablecible con espera	$\lambda$

El índice FES por tramo o usuario se encuentra al sumar las filas de cada columna de la matriz de interrupciones, mientras que el índice por circuito corresponde al índice calculado en tramo del circuito donde se encuentre el equipo de medición. El cálculo del índice FES por tramo se puede realizar aplicando la expresión (6.5).

$$\lambda_{Tn} = \sum_{j=1}^n \lambda_{jn} \quad (6.5)$$

Tabla 6.5: Matriz de interrupciones

Tramo	1	2	...	n
1	$\lambda_{11}$	$\lambda_{12}$	...	$\lambda_{1n}$
2	$\lambda_{21}$	$\lambda_{22}$	...	$\lambda_{2n}$
⋮	⋮	⋮		⋮
n	$\lambda_{n1}$	$\lambda_{n2}$	...	$\lambda_{nn}$
FES	$\lambda_{T1}$	$\lambda_{T2}$	...	$\lambda_{Tn}$

Una vez calculada el número de interrupciones para cada tramo es posible calcular su indisponibilidad anual con las expresiones (6.6) y (6.7).

$$U_{ij} = \lambda_{ij} \cdot r_{ij} \quad (6.6)$$

$$U_{Ti} = \sum_{j=1}^n U_{ij} \quad (6.7)$$

Donde, el subíndice  $ij$  corresponde al número de interrupciones en el tramo  $i$ , debido a una falla en el tramo  $j$ ,  $r_{ij}$  es el tiempo de interrupción por tramo,  $U_{ji}$  es la indisponibilidad anual por tramo y  $UT_i$  es la indisponibilidad anual total por tramo. Estos valores se calculan y reemplazan en la matriz de indisponibilidades mostrada en la tabla 6.6.

Tabla 6.6: Matriz de indisponibilidades

<b>Tramo</b>	<b>1</b>	<b>2</b>	$\dots$	<b>n</b>
<b>1</b>	$U_{11}$	$U_{12}$	$\dots$	$U_{1n}$
<b>2</b>	$U_{21}$	$U_{22}$	$\dots$	$U_{2n}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
<b>n</b>	$U_{n1}$	$U_{n2}$	$\dots$	$U_{nn}$
<b>DES</b>	$U_{T1}$	$U_{T2}$	$\dots$	$U_{Tn}$

El cálculo del índice DES tanto por tramo como por circuito se realiza de la misma forma que para el índice FES. Cabe anotar que se tiene presente que solo existe un equipo de medida que se encuentra localizado en la cabecera del circuito que corresponde a la primer tramo. Si se desea calcular los índices de calidad del servicio por usuario, se debe considerar que estos se encuentran conectados a un tramo, de manera que la frecuencia y duración de las interrupciones corresponde a la del tramo al cual están conectados.

## 6.5. Impacto de la localización de fallas en los índices DES y FES

Se considera el caso más desfavorable en el cual se presenta una falla permanente al final del alimentador (figura 6.2) y los operarios tienen que desplazarse hasta la zona e iniciar el proceso de búsqueda recorriendo el circuito y realizando pruebas para la ubicación del tramo fallado. El recorrido del circuito se realiza a una velocidad promedio de 3 kilómetros por hora y el tiempo que se requiere para realizarlo depende de la longitud del circuito y de la cantidad de ramales que lo conforman.

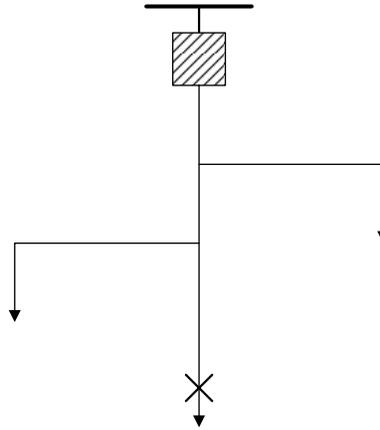


Figura 6.2: Circuito fallado.

Los esquemas más utilizados en la actualidad por las empresas de energía para la identificación y localización de fallas son poco precisos y arrojan como resultado extensas zonas donde se podría ubicar la falla (figura 6.3). Como consecuencia de este hecho es necesario realizar recorridos y pruebas para determinar el lugar exacto de la falla, provocando con esto retardos en la reposición del servicio.

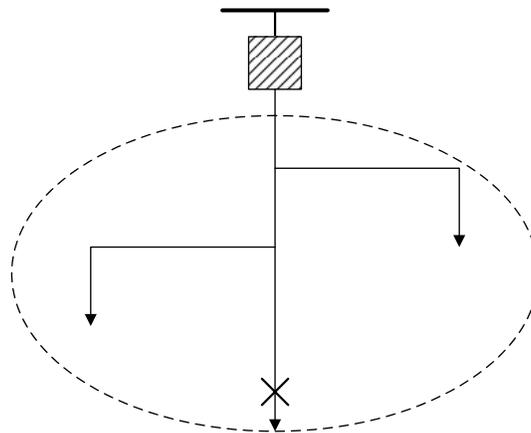


Figura 6.3: Localización de fallas (métodos convencionales).

Los métodos algorítmicos estiman una distancia eléctrica, desde el sitio de medida hasta el lugar de falla, pero tienen el inconveniente de presentar múltiples estimaciones del lugar de la falla debido a la alta ramificación de los circuitos. Este hecho tiene como consecuencia un incremento del tiempo empleado en la localización. Como se observa en la figura 6.4, existe una ubicación posible en cada uno de los ramales del circuito que cumple con esta distancia, en este caso también es necesario recorrer los ramales en los que posiblemente se encuentra la falla, lo cual no representa ninguna disminución considerable en el tiempo de localización.

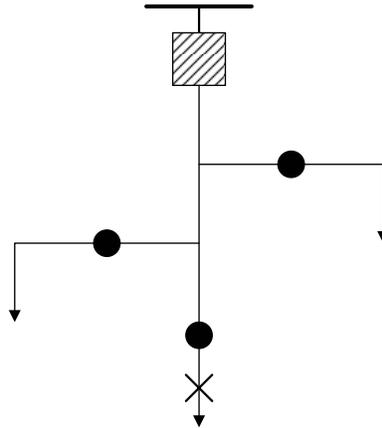


Figura 6.4: Localización de fallas (métodos algorítmicos).

Las técnicas de clasificación eliminan el problema de múltiple estimación, pero dependen de la calidad de los datos suministrados por el sistema de información de la empresa. La precisión no es tan alta como la de los métodos algorítmicos. En este caso se reduce el tamaño de la zona de búsqueda y se reduce un poco el tiempo de localización (figura 6.5).

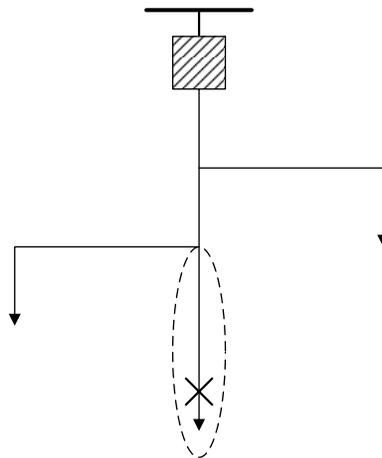


Figura 6.5: Localización de fallas (técnicas de clasificación).

Con la unión de estas dos técnicas se puede obtener una mejor estimación, debido a que estas se complementan mutuamente. Es así como interceptando los posibles puntos de fallas arrojados por los métodos algorítmicos con la zona de falla obtenida con las técnicas clasificación se logra una respuesta de localización mucho más precisa que las anteriores (figura 6.6). La intercepción entre los resultados arrojados por las dos técnicas permite reducir el tiempo empleado en el recorrido de todo el circuito para encontrar el punto con falla y sólo es necesario acceder a él y

determinar el dispositivo que ha fallado.

Con esta nueva información de localización es posible reducir considerable el tiempo de restauración del suministro después de la reparación de la falla y por lo tanto el mejoramiento de los índices de calidad DES y FES. Debido a que la implementación de estos métodos afectan directamente los intervalos  $T_e$  y  $T_l$  disminuyéndolos al no ser necesario realizar contraste de información entre la base de datos de la empresa y la información proporcionada por los usuarios. Además se hace innecesario la ejecución de pruebas y recorridos extensos para localizar la falla.

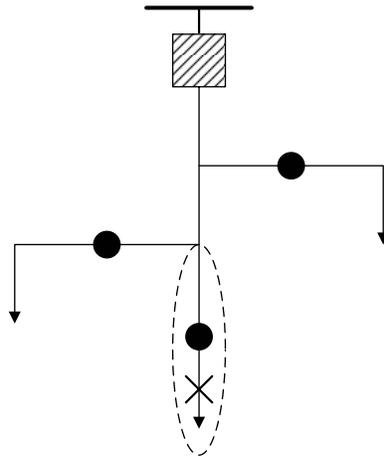


Figura 6.6: Localización de fallas (métodos algorítmicos y técnicas de clasificación).



# Capítulo 7

## Conclusiones y trabajos futuros

### 7.1. Conclusiones

El autor expresa su total agrado y satisfacción hacia el grupo de trabajo con el cual se culmina en buenos terminos el desarrollo de este proyecto, el cual, contribuye al fortalecimiento de la línea de investigación en *Calidad del Servicio de Energía Eléctrica* del grupo de investigación GISEL.

En este trabajo de investigación se ha tratado la aplicación de técnicas de la estadística multivariada a la localización de fallas en sistemas de distribución, destacando los resultados obtenidos al utilizar una pequeña cantidad de descriptores durante la construcción del modelo y la precisión obtenida.

Este trabajo contribuye a satisfacer la necesidad actual de las empresas distribuidoras y operadores de red por mantener y mejorar la calidad del servicio, a partir de la disminución de la duración y el número de las interrupciones.

Se presenta una novedosa alternativa de solución al problema de localización de fallas en los sistemas de distribución, la cual, está compuesta por una herramienta estadísticamente soportada, basada en la exploración y modelado del conocimiento almacenado en la información del sistema, cuando se presenta una falla.

El modelo es el resultado de la implementación de simulaciones del sistema bajo condiciones de falla y la aplicación de técnicas estadísticas a la información obtenida. El empleo de estas técnicas facilita la interpretación de los datos y permite la conformación de grupos asociados a zonas del sistema. El algoritmo utilizado para realizar los agrupamientos posee valores iniciales de

operación, los cuales son responsabilidad del analista, debido al conocimiento previo y profundo de los datos que van a ser procesados.

Un factor influyente para el modelo es la detección e identificación del tipo de falla, debido a que en la propuesta se trabaja cada uno de los tipos de falla de forma independiente. De tal forma que los grupos obtenidos se puedan identificar visualmente, sin embargo, se observa que no siempre se obtiene esta clase de resultados.

El desempeño del modelo depende de la calidad en la obtención y procesamiento de la información (sistemas de información confiables), además de la caracterización detallada del sistema al cual se aplica. Esto conlleva la necesidad de poseer sistemas de información confiables por parte de las empresas.

Los requerimientos de software para la implementación de la herramienta propuesta son suministrados por paquetes estadísticos de uso comercial y de bajo costo, lo cual permite establecer la sencillez y facilidad en la implementación del modelo. Los requerimientos de hardware son básicos y económicos, ya que depende únicamente de un sistema de adquisición de datos que esté en capacidad de monitorear las señales de tensión y corriente en la cabecera del circuito.

El modelo propuesto contempla las implicaciones, técnicas, económicas y operativas existentes en los sistemas de distribución, destacando la mínima inversión en la implementación de la herramienta adecuada por parte de las empresas para la prevención y atención oportuna de fallas.

Existen limitaciones potenciales para el modelo propuesto, como lo son, la selección de la cantidad de grupos a conformar, la proporción de los grupos en la muestra y los valores de inicio de los algoritmos utilizados. Para superar estas limitaciones se utilizaron criterios teóricos y heurísticos, siendo los últimos los de mayor influencia en la estructura del modelo ya que hacen parte del conocimiento a priori que posee el analista sobre la información.

Se realiza una propuesta metodológica a partir del fraccionamiento del tiempo de indisponibilidad debido a una falla, en tiempos asociados a las acciones realizadas durante la atención de la misma y posterior recuperación del sistema.

Al desagregar el tiempo total de una interrupción en una serie de intervalos, los cuales tienen asociadas actividades específicas, se observa que cada intervalo juega un rol importante en el

proceso de recuperación de la falla y restauración de la operación del sistema. El empleo de los datos de localización afecta directamente los intervalos de tiempo asociados a la estimación del lugar, el traslado y la localización de la falla, disminuyendo de forma apreciable el tiempo total de una interrupción.

El desarrollo detallado de las actividades realizadas en cada intervalo de tiempo durante una interrupción, es una herramienta que puede aplicarse a la mayoría de las empresas de distribución del sector. Las actividades asociadas a cada intervalo y la duración de las mismas dependen de los recursos técnicos y humanos con los que cuente la empresa.

Lograr una rápida recuperación de fallas y restauración de la operación del sistema con el objeto de mejorar los índices de calidad del servicio, depende de la frecuencia y de la duración de las interrupciones. Sin embargo, se observa que es más conveniente intervenir en la disminución de la duración que de la frecuencia, debido a la naturaleza aleatoria que presentan las fallas.

Este proyecto fortalece la cultura de la calidad del servicio generando a mediano plazo, beneficios económicos a las empresas y mayor grado de satisfacción a los usuarios.

## 7.2. Aportes

Se diseñó una herramienta software para la construcción, validación y evaluación del modelo estadístico para la localización de fallas en sistemas de distribución, en Matlab.

Se plantea una metodología para el diseño y la construcción del modelo. La herramienta utiliza algoritmos de fácil consulta y dominio publico.

Se incursiona en la aplicación de las mezclas de distribuciones y sus algoritmos auxiliares en la búsqueda de alternativas de solución a problemas de actualidad relacionadas con la ingeniería eléctrica.

Se implementó una herramienta software en MatLab, para la localización de fallas basada en el método algorítmico Ratan Das. El método realiza la estimación de la sección posible de falla, además de la distancia entre el punto de falla y el punto de medida. Destacando que el número de estimaciones depende de la topología del sistema de distribución.

Se implementó una herramienta software en MatLab, para el reconocimiento de patrones que no se limita solamente al análisis y solución del problema de la localización de fallas. Es decir, la herramienta está diseñada puede ser utilizado para cualquier tipo de problema multivariado.

Se ha elaborado una metodología para evaluar la incidencia de los datos de localización en el tiempo total de interrupción, y estimar los índices de calidad (DES y FES) durante una indisponibilidad en sistemas de distribución que operan en forma radial.

### 7.3. Trabajos Futuros

El modelo propuesto deja al descubierto potenciales relaciones entre los datos, tal y como se muestra en los resultados presentados. Es de notar que el nivel de detalle del modelo puede aumentar con el incremento del número de descriptores, no obstante, estos nuevos descriptores sean fáciles de calcular que relacionen adecuadamente la información, debido a que se podría incurrir en el aumento de la complejidad del modelo. Por lo tanto, la relación entre la precisión y la complejidad del modelo es un criterio a tener presente en todo momento por el analista de los datos, dadas las características actuales de los equipos de cómputo.

Se promueve la concepción de estrategias enfocadas hacia la eficiencia en el reestablecimiento de las condiciones de operación del sistema, a partir de la mejora en los tiempos de atención y recuperación de fallas, además de la prevención y corrección de fallas permanentes del sistema.

Se propone como etapa siguiente del proyecto, la evaluación del desempeño del modelo con información obtenida a partir de sistemas de distribución reales.

Se deja como precedente la opción de incluir construir un modelo que mezcla las mejores características de un método algorítmico con las del modelo estadístico.

La herramienta solo maneja variables cuantitativas, queda abierto el campo para incluir en éste análisis la caracterización por medio de variables cualitativas, a partir de la aplicación de técnicas apropiadas para este fin.

Se propone incentivar el trabajo de asimilación de otras técnicas de clasificación reconocidas, con el objeto de afianzar el conocimiento en esta área y proponer aplicaciones similares en otras

áreas de la ingeniería eléctrica.

Esta metodología se podrá validar a partir de la toma de tiempos en las actividades planteadas por parte de las empresas de distribución, teniendo presente información acerca de la topología del sistema, tasas de falla y causología. De igual manera se puede realizar un manual de procedimientos en donde se den algunos parámetros importantes para los procedimientos de atención y recuperación de fallas.

Con el apoyo de esta metodología se podrán realizar estudios sobre la implementación de nuevas tecnologías en los sistemas de distribución, utilizando como criterio de selección la estimación de los valores de los índices de calidad, con el objeto de justificar la inversión.



# Apéndice A

## Parámetros del sistema de distribución de prueba

Los parámetros del sistema de distribución utilizado para desarrollar la metodología planteada en el capítulo 4 se presentan a continuación. En el modelo de la figura A.1 fueron realizadas las simulaciones de las fallas de cortocircuito.

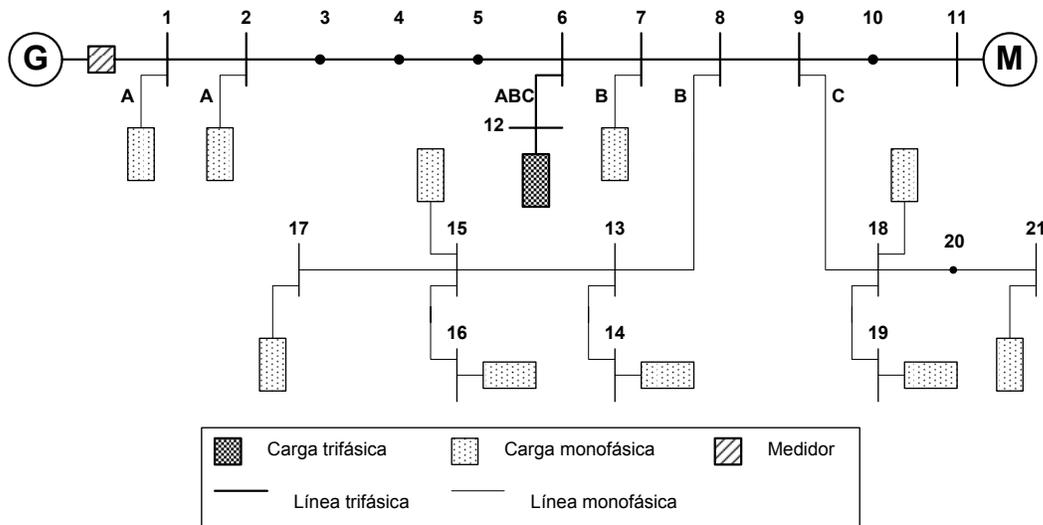


Figura A.1: Diagrama unifilar del sistema de distribución.

Los parámetros de la fuente ubicada en la cabecera del circuito se muestran en la tabla A.1, así como la información de cada una de las cargas ubicadas en cada una de las barras se muestran en la tabla A.2. La descripción de los parámetros de las líneas que interconectan las barras se muestran en la tabla A.3 y por último el factor de potencia de las cargas utilizadas están definidos la tabla A.4.

Tabla A.1: Parámetros del generador.

Tensión Base [kV]	Potencia Base [MVA]	$Z_+$ y $Z_-$ [p.u]	$Z_0$ [p.u]
25	100	$0,68283 + j2,98139$	$0,09496 + j1,39289$

Tabla A.2: Parámetros de las de cargas del sistema.

Nodo	Fase	Carga [kVA]	Composición [%]		
			Calentador	Iluminación	Motor
1	A	15	99,8	0,1	0,1
2	A	15	99,8	0,1	0,1
7	B	15	99,8	0,1	0,1
11	A, B, C	1000	0,1	0,1	99,8
12	A, B, C	67,5	99,8	0,1	0,1
14	B	15	99,8	0,1	0,1
15	B	15	99,8	0,1	0,1
16	B	7,5	99,8	0,1	0,1
17	B	15	99,8	0,1	0,1
18	C	25	99,8	0,1	0,1
19	C	15	99,8	0,1	0,1
21	C	15	99,8	0,1	0,1

Los datos de la línea se presentan por secciones. Los nodos a los que está conectada cada sección, la distancia de cada sección y las impedancias del modelo equivalente  $\pi$  se muestran en la tabla A.3. Algunas características del sistema propuesto son:

- Entre los nodos 1 y 11 existen 37km de longitud. Esta línea esta compuesta por diferentes tipos de conductores.
- Se instalaron cargas de diferentes tipos en los nodos del sistema excepto en los nodos 3, 4, 5, 10 y 20.
- Los nodos 3, 4 y 5 forman parte de una sección de 16km.
- El nodo 20 divide en dos secciones la sección entre los nodos 18 y 21.
- El nodo 10 es la unión de dos secciones de diferente tipo de conductor.

Tabla A.3: Parámetros de las líneas

Sección entre nodos	Longitud [km]	$Z_{\text{Serie}}$ [ $\Omega/\text{km}$ ]		$Y_{\text{Paralelo}}$ [Mhos/km]	
		$Z_+$ y $Z_-$	$Z_0$	$Y_+$ y $Y_-$	$Y_0$
1 - 2	2,414	0,3480 + j0,5166	0,5254 + j1,704	j3,74e-6	j2,49e-6
2 - 6	16,092	0,3480 + j0,5166	0,5254 + j1,704	j3,74e-6	j2,49e-6
6 - 7	4,023	0,3480 + j0,5166	0,5254 + j1,704	j3,74e-6	j2,49e-6
7 - 8	5,150	0,5519 + j0,5390	0,7290 + j1,727	j3,59e-6	j2,39e-6
8 - 9	2,414	0,5519 + j0,5390	0,7290 + j1,727	j3,59e-6	j2,39e-6
9 - 10	4,506	0,5519 + j0,5390	0,7290 + j1,727	j3,59e-6	j2,39e-6
10 - 11	2,414	0,3480 + j0,5166	0,5254 + j1,704	j3,74e-6	j2,49e-6
11 - 12	2,414	0,3480 + j0,5166	0,5254 + j1,704	j3,74e-6	j2,49e-6
8 - 13	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
13 - 14	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
13 - 15	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
15 - 16	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
15 - 17	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
9 - 18	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
18 - 19	2,414	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
18 - 20	3,219	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6
20 - 21	3,219	7,3977 + j0,8998	7,3977 + j0,8998	j2,51e-6	j2,51e-6

Tabla A.4: Factor de potencia de las cargas.

Tipo de carga	FP
Calentador	1
Iluminación	0,85 atraso
Motor	0,8 atraso



# Bibliografía

- Aguilar J. y Aguado C. "A mixed qualitative quantitative self learning classification technique applied to situation assessment in industrial control. 1999.
- Anderson P. *Analysis of faulted power systems*. IEEE Press, New York, 1995.
- Arriaga A. "Evaluación de Confiabilidad en Sistemas Eléctricos de Distribución". Master's thesis, Universidad Católica de Chile, Santiago de Chile, 1994.
- Arrillaga J. *Power system quality assessment*. Wiley, New Zeland, 2000.
- Aucoin B. y Jones R. "High Impedance Fault Implementation Issues". *IEEE Transactions on Power Delivery*, Vol. 11, 1996.
- Barrera V. "Sistemas de distribución: metodología para localización de fallas mediante la aplicación de inteligencia artificial". Master's thesis, Universidad Industrial de Santander, Colombia, 2006.
- Beltrán D. y Romero A. "Calidad del servicio: propuesta metodológica para la estimación de los índices de calidad del servicio DES y FES a partir de información de localización de fallas". 2006.
- Bollen M. *Understanding power quality problems: voltages sags and interruptions*. IEEE Press, New York, 2000.
- Brown R. *Electric power distribution reliability*. Marcel Dekker, New York, 2002.
- Carrillo G. y Ordóñez G. "Calidad del Suministro de Energía Eléctrica (Power Quality)". *ALURE Program - CREG Project Tutorial Course*. 2002.
- Carrillo G., Román J. y Rivier J. "Metodología coste beneficio aplicada las instalaciones de equipos de localización seccionamiento del tramo con avería en redes de distribución". *Actas de las 3ras jornadas hispanolusas de ingeniería eléctrica*. 1993.

- CIDET . *Reglamento Técnico de Instalaciones Eléctricas-RETIE*. Corporación y centro de investigación desarrollo tecnológico <http://www.cidet.com.co/>, 2005.
- Cormane J., Barrera V. y Ordonez G. “Predicción de corrientes armónicas en ventanas de carga residenciales mediante modelos neuronales artificiales”. *II Congreso Internacional de la Región Andina (Andescon-IEEE)*. 2004.
- Cormane J., Barrera V., Rodríguez J., Carrillo G., Ordóñez G., Vargas H. y Mora J. “Modelos Híbridos para la Localización de Fallas en Sistemas de Distribución de Energía Eléctrica”. Technical report, Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones-UIS, Colombia, 2005.
- Cárdenas M. y Olaya O. “Norma para las actividades del área atención inmediata en empresas distribuidoras de energía eléctrica en Colombia”. 2002.
- CREG . “Concepto técnico sobre procedimientos operativos, tiempos admisibles y frecuencia en la ejecución de mantenimientos programados”. Technical report, Comisión Reguladora de Energía y Gas-CREG, Bogotá, 2004.
- CREG . *Resolución CREG 070/1998, CREG 096/2000, CREG 084/2002, CREG 025/2005*. Comisión de Regulación de Energía y Gas <http://www.creg.gov.co/>, 2006.
- Dagenhart J. “The 40-Ground-Fault Phenomenon”. *IEEE Transactions on industry applications*, Vol. 36, 2000.
- Das R. *Determining the Locations of Faults in Distribution Systems*. PhD thesis, Saskatchewan University, Canada, 1998.
- de Sá J. M. *Pattern Recognition : Concepts, Methods and Applications*. Springer Verlag, New York, 2001.
- Diaz L. *Estadística Multivariada: Inferencia y Métodos*. UNAL, Bogotá, 2002.
- Duda R., Hart P. y Stork J D. *Pattern Classification*. Wiley Interscience, New York, 2000.
- Figueiredo M. “Unsupervised learning of finite mixture models”. *IEEE Transactions pattern analysis and machine intelligence*, Vol. 24, 2002.
- Girgis A., Fallon C. y Lubkeman D. “A fault location technique for rural distribution feeders”. *IEEE Transactions Industry Applications*, Vol. 29, 1993.

- Grainger J. y Stevenson W. *Análisis de sistemas eléctricos de potencia*. McGraw Hill, Mexico, 1999.
- Greenwood A. *Electrical Transients in Power Systems*. Wiley, New York, 1991.
- Hair J., Anderson R., Tatham R. y Black W. *Multivariable Data Analysis*. Prentice Hall, New York, 1999.
- IEEEstd1159-1995. "IEEE Recommended practice for monitoring electric power quality". 1995.
- Isaza C. "Clasificación y diagnóstico de procesos empleando metodología LAMDA". Master's thesis, Universidad de los Andes, Colombia, 2004.
- Jaimes J. "Conceptualización técnica del servicio de reserva operativa". Master's thesis, Universidad Industrial de Santander, Colombia, 2003.
- Jaimes J., Duarte J., Ordóñez G., Carrillo G. y Gaona P. "La calidad de la energía eléctrica en el sector eléctrico colombiano: utopía o realidad". *II Congreso Internacional de la Región Andina (Andecon-IEEE)*. 2004.
- Lee S., Choi M., Kang S., Jin B., Lee D., Ahn B., Yoon N., Kim H. y Wee S. "An intelligent and efficient fault location and diagnosis scheme for radial distribution systems". *IEEE Transactions on Power Delivery*, Vol. 19, 2004.
- López L. "Localización de fallas: Herramienta de clasificación basada en mezclas finitas". 2006.
- Martínez W. y Martínez A. *Computational statistics Handbook whit MatLab*. Chapman Hall, New York, 2002.
- Mclachlan G. y Peel D. *Finite mixture models*. Wiley, Canada, 2000.
- Meléndez J., Mora J. y Carrillo G. "Una arquitectura genérica para el desarrollo de sistemas híbridos. Propuesta de localizador de fallas para mejorar los índices de calidad". *III Simposio Internacional sobre Calidad de la Energía Eléctrica SICEL 2005*. 2005.
- Mora J. y Colomer J. "Voltage Sag Characterization and Classification for Diagnosis in Electric Power Quality Domain". Technical report, Department of Electronics, Computer Science and Automatic Control-UDG, España, 2003.
- Mora J., Serrano H. y Ordóñez G. "Mejoramiento de los índices de calidad del servicio mediante la caracterización de señales para localización de fallas en distribución". *III Simposio Internacional sobre Calidad de la Energía Eléctrica SICEL 2005*. 2005.

- Morales G. y Gómez A. "Estudio e implementación de una herramienta basada en máquinas de soporte vectorial para la localización de fallas en sistemas de distribución". 2005.
- Novosel D., Hart D., Hu Y. y Myllymaki J. "System for locating faults and estimating fault resistance in distribution networks with tapped loads". 1998.
- Oppenheim A. y Willsky A. *Señales y sistemas*. Prentice Hall, México, 1998.
- Peña D. *Análisis de datos multivariantes*. McGraw Hill, Madrid, 2002.
- Peñuela A., Pacheco J., Ordóñez G. y Duarte J. "Propuesta metodología para la valoración de costos en empresas manufactureras, debidos a una inadecuada continuidad del suministro de la energía eléctrica". 2004.
- Rencher A. *Methods of Multivariate Analysis*. Wiley Interscience, New York, 1995.
- Rodríguez J. "Detección y localización de fallas en los sistemas de energía eléctrica mediante la técnica máquinas de soporte vectorial". Master's thesis, Universidad Industrial de Santander, Colombia, 2006.
- Suárez M. y Salamanca E. "Localización de fallas: reconocimiento estadístico de patrones". 2006.
- Uriarte F. "Modeling, detection and localization of high-Impedance faults in low voltage distribution feeders". Master's thesis, Virginia Tech Polytechnic Institute, Estados Unidos, 2003.
- Vargas H. y Cormane J. "Convenio interadministrativo entre la Electrificadora de Santander y la UIS para la detección de consumos anómalos de energía eléctrica". Technical report, Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones-UIS, Colombia, 2004.
- Vega V., C. Duarte y Ordóñez G. "Detección e identificación de eventos de la calidad de la energía eléctrica utilizando la Transformada Wavelet Discreta y redes neuronales artificiales". *III Simposio Internacional sobre Calidad de la Energía Eléctrica SICEL 2005*. 2005.
- Villamizar L. y Quiñones C. "Implementación del método Ratan Das para la localización de fallas en sistemas de distribución de energía eléctrica". 2005.
- Wang Y., Lin S., Li H. y S.Kung. "Data mapping by probabilistic modular networks and information theoretic criteria". *IEEE Transactions Signal processing*, Vol. 46, 1998.
- Wang Y., Luo L., Freedman M. y Kung S. "Probabilistic principal component subspaces: A hierarchical finite mixture model for data visualization". *IEEE Transactions Neural Networks*, Vol. 11, 2000.

Zhu J., Lubkeman D. y Girgis A. "Automated fault location and diagnosis on electric power distribution feeders". *IEEE Transactions Power Delivery*, Vol. 12, 1997.